

OM of: US-09-303-518d-127 to: SwissProt_40:* out_format: pfs
Date: Jun 30, 2002 8:33 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=firemap_nzp_model -DEV=xlh
-Q=cn2_1/USPPOI/US09303518/runat_28062002_147474.4362/app_query.fasta_1.23501
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DEL0P=6.000 -DEL0P=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DEL0P=6.000 -DEL0P=7.000 -START=1 -MATRIX=blsnum62
-TRAN=human40.cdi -LIST=100 -DOCALLIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=5 -MODE=LOCAL -OUTFILE=pfs
-NORMEXT -HARSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09303518 -ECGN1_1.440 -NCPD=6 -ICPD=3 -LONCLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -NO_XLPHYX -WAIT -THREADS=1

Search information block:

Query: US-09-303-518d-127
Query length: 1344
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000

Score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SwissProt_40:NORA_NEIMA + 2211.00 3356.88	2.3e-174	447	0.0174	447	091y08 neisseria meningitidis
SwissProt_40:NORA_NEIMA + 2177.00 3306.60	1.5e-171	447	0.0171	447	091y08 neisseria meningitidis
SwissProt_40:NORA_PASIM + 1614.50 2374.79	3.1e-125	446	0.0125	446	091y08 neisseria meningitidis
SwissProt_40:NORA_HAEIN + 1574.00 2314.87	6.8e-122	446	0.0122	446	091y08 neisseria meningitidis
SwissProt_40:NORA_VIBAL + 1424.50 2090.85	1.4e-109	446	0.0109	446	091y08 neisseria meningitidis
SwissProt_40:NORA_VIBCH + 1412.50 2076.06	1.4e-108	446	0.0108	446	091y08 neisseria meningitidis
SwissProt_40:NORA_PSEAE + 1298.00 1906.76	3.7e-99	445	0.0099	445	091y08 neisseria meningitidis
SwissProt_40:NORA_CHIMU + 447.50 639.57	4.2e-29	467	0.0029	467	092742 chlamydia pneumoniae
SwissProt_40:NORA_CHLTR + 445.00 615.34	3.0e-28	465	0.0028	465	092742 chlamydia pneumoniae
SwissProt_40:NRNC_YIBCH + 157.00 214.17	4.0e-05	774	0.0005	774	091y08 neisseria meningitidis
SwissProt_40:NRNC_YIBCH + 136.00 175.08	0.0026	1802	0.0026	1802	091y08 neisseria meningitidis
SwissProt_40:NRNC_YIBCH + 131.00 172.04	0.0061	1140	0.0061	1140	091y08 neisseria meningitidis
SwissProt_40:NRNC_YIBCH + 123.50 160.47	0.0255	1199	0.0255	1199	091y08 neisseria meningitidis
SwissProt_40:NRNC_YIBCH + 123.00 158.48	0.0288	1367	0.0288	1367	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 121.00 150.93	0.0372	774	0.0372	774	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 116.50 154.00	0.0878	797	0.0878	797	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 116.50 153.74	0.0884	819	0.0884	819	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 110.00 141.18	0.3245	1117	0.3245	1117	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 109.50 148.94	0.2933	457	0.2933	457	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 108.00 147.73	0.3016	539	0.3016	539	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 107.50 147.83	0.4106	437	0.4106	437	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 107.50 141.59	0.4742	725	0.4742	725	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 107.50 145.83	0.5854	1229	0.5854	1229	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 106.50 134.01	0.5239	481	0.5239	481	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 105.50 131.05	0.8254	1609	0.8254	1609	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 105.00 133.51	0.8430	1150	0.8430	1150	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 103.50 125.69	1.128	2073	1.128	2073	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 103.00 139.79	0.9953	435	0.9953	435	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 103.00 138.84	1.102	481	1.102	481	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 103.00 134.74	1.112	740	1.112	740	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 101.00 132.64	1.160	676	1.160	676	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 101.00 127.50	1.80	1161	1.80	1161	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 100.50 136.64	1.58	411	1.58	411	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 100.50 135.30	1.63	473	1.63	473	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 100.50 121.43	2.24	2035	2.24	2035	091y08 neisseria meningitidis
SwissProt_40:NRNC_PSEAE + 100.00 135.29	1.76	438	1.76	438	091y08 neisseria meningitidis

SwissProt_40:PMAL_DUNAC + 100.00 126.51	2.16	1103	2.16	1103	p54210 dunaliella acidop
SwissProt_40:YGLF_YEAST + 99.50 132.37	2.04	551	2.04	551	p53214 saccharomyces cere
SwissProt_40:YGLF_YEAST + 99.00 118.96	3.00	2090	3.00	2090	p51611 mesocricetus astra
SwissProt_40:YGLF_YEAST + 98.50 127.39	2.59	695	2.59	695	094477 mus musculus (mou
SwissProt_40:YGLF_YEAST + 98.50 127.39	2.59	796	2.59	796	090625 sus scrofa (pig). fo
SwissProt_40:YGLF_YEAST + 98.00 121.95	3.27	1306	3.27	1306	p33234 saccharomyces ele
SwissProt_40:YGLF_YEAST + 98.00 116.06	3.75	2426	3.75	2426	p33234 saccharomyces ele
SwissProt_40:YGLF_YEAST + 97.50 126.52	3.04	605	3.04	605	p18583 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 97.50 126.23	3.20	770	3.20	770	p18583 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 97.50 126.16	3.21	776	3.21	776	p18583 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 97.00 122.98	3.73	1003	3.73	1003	p40562 azobacterium vinea
SwissProt_40:YGLF_YEAST + 96.50 131.27	3.33	388	3.33	388	p18287 g trifunctional p
SwissProt_40:YGLF_YEAST + 96.50 122.92	4.36	864	4.36	864	p52424 vigne unguiculata
SwissProt_40:YGLF_YEAST + 95.50 126.37	4.36	556	4.36	556	09142 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 95.00 122.15	5.19	802	5.19	802	09142 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 94.50 124.47	5.32	581	5.32	581	029101 mus musculus (mou
SwissProt_40:YGLF_YEAST + 94.50 123.70	5.41	630	5.41	630	002496 mus musculus (mou
SwissProt_40:YGLF_YEAST + 94.00 122.03	6.08	695	6.08	695	p32212 caenorhabditis ele
SwissProt_40:YGLF_YEAST + 93.50 124.01	6.28	522	6.28	522	027245 caenorhabditis ele
SwissProt_40:YGLF_YEAST + 93.00 123.49	6.87	1251	6.87	1251	009550 caenorhabditis ele
SwissProt_40:YGLF_YEAST + 93.00 119.61	7.51	510	7.51	510	p07001 escherichia coli.
SwissProt_40:YGLF_YEAST + 93.00 112.69	8.81	819	8.81	819	p04065 saccharomyces dias
SwissProt_40:YGLF_YEAST + 92.50 115.99	8.05	1589	8.05	1589	092427 pyrococcus abyssi
SwissProt_40:YGLF_YEAST + 92.50 115.06	9.02	682	9.02	682	p50478 gallus gallus (chi
SwissProt_40:YGLF_YEAST + 92.50 114.15	9.21	1260	9.21	1260	p24384 saccharomyces cer
SwissProt_40:YGLF_YEAST + 92.00 123.48	8.03	437	8.03	437	091420 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 92.00 122.42	8.29	440	8.29	440	p63172 rhodococcus erythr
SwissProt_40:YGLF_YEAST + 92.00 119.04	8.89	507	8.89	507	053743 zymomonas mobilis
SwissProt_40:YGLF_YEAST + 92.00 114.27	9.93	697	9.93	697	p17886 mycobacterium tube
SwissProt_40:YGLF_YEAST + 92.00 114.00	9.99	1152	9.99	1152	p27816 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 91.50 119.73	9.46	600	9.46	600	p23226 dirosophila melano
SwissProt_40:YGLF_YEAST + 91.50 114.88	10.58	999	10.58	999	p14328 dirosophila melano
SwissProt_40:YGLF_YEAST + 91.00 115.28	11.33	886	11.33	886	p18596 rattus norvegicus (r
SwissProt_40:YGLF_YEAST + 91.00 115.06	11.39	907	11.39	907	p00284 epstein-barr virus
SwissProt_40:YGLF_YEAST + 90.50 115.90	12.07	768	12.07	768	p03200 epstein-barr virus
SwissProt_40:YGLF_YEAST + 90.00 122.16	11.29	368	11.29	368	p24384 saccharomyces cer
SwissProt_40:YGLF_YEAST + 90.00 120.92	11.62	419	11.62	419	091420 homo sapiens (huma
SwissProt_40:YGLF_YEAST + 90.00 119.89	11.90	467	11.90	467	p50770 human papillomavir
SwissProt_40:YGLF_YEAST + 90.00 116.87	12.76	642	12.76	642	p32599 saccharomyces cere
SwissProt_40:YGLF_YEAST + 89.50 114.58	13.45	817	13.45	817	p42850 pyrococcus furiosu
SwissProt_40:YGLF_YEAST + 89.50 121.68	12.34	358	12.34	358	p39812 b aroa(g) protein
SwissProt_40:YGLF_YEAST + 89.50 119.17	13.08	466	13.08	466	p54208 synecococcus sp.
SwissProt_40:YGLF_YEAST + 89.50 117.43	13.62	466	13.62	466	p50801 human papillomavir
SwissProt_40:YGLF_YEAST + 89.50 116.94	13.80	560	13.80	560	p00589 corynebacter beta.
SwissProt_40:YGLF_YEAST + 89.50 116.84	14.42	596	14.42	596	p53379 saccharomyces cere
SwissProt_40:YGLF_YEAST + 89.00 117.80	14.59	728	14.59	728	092512 streptomyces coeli
SwissProt_40:YGLF_YEAST + 89.00 117.80	14.59	498	14.59	498	p26529 nicotiana glumigil
SwissProt_40:YGLF_YEAST + 88.50 116.88	14.90	498	14.90	498	p26530 nicotiana glumigil
SwissProt_40:YGLF_YEAST + 88.50 117.70	15.81	549	15.81	549	p15337 escherichia coli.
SwissProt_40:YGLF_YEAST + 88.50 112.96	16.31	829	16.31	829	p44323 haemophilus influen
SwissProt_40:YGLF_YEAST + 88.50 117.22	15.98	466	15.98	466	p17389 human papillomavir
SwissProt_40:YGLF_YEAST + 88.50 113.31	17.19	490	17.19	490	p31060 escherichia coli.
SwissProt_40:YGLF_YEAST + 88.50 108.48	19.55	739	19.55	739	p15143 gallus gallus (chi
SwissProt_40:YGLF_YEAST + 88.50 108.38	19.55	1228	19.55	1228	p35825 bacillus stearoth
SwissProt_40:YGLF_YEAST + 88.50 96.29	25.90	4427	25.90	4427	p005470 bacillus subtilis

seq_name: SwissProt_40:NORA_NEIMA

seq_documentation_block:

ID NORA_NEIMA STANDARD; PRT; 447 AA.

AC Q9VPE6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Na(+)-translocating NADH:quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A) (NQR complex
subunit A) (NQR1 subunit A).
GN NORA OR NMA0752.

OS *Neisseria meningitidis* (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 NCBI_TaxID=65699;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jags K., Leather S., Moul S., Mungall K., Quail M.A., Ralstead M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.";
 RT Nature 404:502-506(2000).
 CC -1- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORA ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF UBISEMQUINONE TO UBIQUINOL (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) + UBIQUINOL + NA(+)(OUT).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE AND NORE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: AL162754; CAB84035.1;
 CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 KW Complete proteome.
 SQ SEQUENCE 447 AA; 48673 MW; B808CAA82C901AD CRC64;

alignment_scores:
 Quality: 2211.00 Length: 447
 Ratio: 5.059 Gaps: 0
 Percent Similarity: 97.763 Percent Identity: 97.092

alignment_block:

US-09-303-518D-127 x NORA_NEIMA

Align seg 1/1 to: NORA_NEIMA from: 1 to: 447

```

1 ATGATTAATCAAAAAAGCTTAACCTGCCCTCCGGGCGACCGGA 50.
|||||
1 Mettleysileyslysglyleuansleurollealaglyagprogl 17
51 GCAAGTCATTATGACGGCGCCGTCATTACGAGAGTGGCGTGGCG 100
|||||
17 uginvalilletyraspelyprovalillethngluvalalaleuenclyg 34
101 AAGAATATGCCGTATGCGCCCTNGATGAAGTCAAGAGAGCGATGCC 150
|||||
34 luguityralaglymetarprosermetelyslysgluylaspala 50
151 GTCAAAAAAGGCCAAGTCTGTGAGACAAAAAGNATCCGGCGCGT 200
|||||
51 Vallyslysglygluvalleuhegluasplyslysanproglyvala 67
201 GTTACCGCGCGCGTTCAGGCAAAATCGCGCATCCATCGCGCGAAA 250
|||||
67 lphetrlalprovalserglylysilealalalehisargglyglul 84
251 AGCGGTAATTCACTGCTGCTGATGCGCTGAAGCAACGACGAATC 300
|||||

```

```

84 ySArgValleuGlnSerValIlealValgluylasnaspsluile 100
301 GAGTTCGACCGCTACGGCGCGGAGCGTTGGCAAACTTAACGGCGANGA 350
|||||
101 GUphegluaArgylalalprogluValleuValasneuserglyglul 117
351 ATNNNGNNGCAATGATTCATCCGATCGGTTTGGACTGCGTGGTANCC 400
|||||
117 uValalgarValasneuleuIleGlnserglyLeuTrpThalaleuArgTha 134
401 GTCCGTTAGCAAAATCCCTGCGTCGATGCCGAGCGGTGGCATCTTC 450
|||||
134 rgrProPheSerlyslleProalalaspalalagluProPhealallephe 150
451 GTCAATGGATGACCAACATCCGCTNCGCGCAGACCTGTGGTGTGAT 500
|||||
151 ValasnalameValasprThasProleuValalalaspProvalalVal 167
501 CAAGAAGCGCGANGATTCAGACGANGTNGCTGTATTCAGCGCTT 550
|||||
167 elysgluValalagluasprhealargglyleuValleuSerargl 184
551 TCACCGAGCGTAAATCCATGTGTGTAGCGACGTCGGCGCAGAGTGGC 600
|||||
184 eutThrgluArglysllehisvalcyslysalalaglyalaspvalpro 200
601 TCTGAATATGCTGCCACATCGAACACATGATTCGGCGCGCGCATCC 650
|||||
201 SergluasalalalalasnilleluThrhsgluPheglylprohisPr 217
651 GCGCGGTTTGAAGTGGCAGCGACATTCATTTCATTCAGCGCGTGGTGA 700
|||||
217 oAlaglyleuSerglyThrhsllehisPheillegluProValglyalaa 234
701 ACAAACGCTTGGACACATCATTCATCAAGATGATTCGATGGAGAGT 750
|||||
234 snlysthrValtrpThrhleasnytrgluasvalillealillelytrg 250
751 TTGTTGCAACAGCGCGTGTGACACGAGCGCGATTCGTTGGTGGTGG 800
|||||
251 leuphealathrclylArgleuasnhrcluarvalillealaleuclgyl 267
801 TTCTCAAGTCAACAAACCAACGCGCTTCGTAACGCTTTGGTGGCGAAG 850
|||||
267 ysergluValasnllyspProargleuLeuargThrValleuGlyalalsv 284
851 TATCGCAATTTACGCGCGCAATTCGTTGACGACGACGCGCGTGAAT 900
|||||
284 alSerGlnleThrhAlaglygluValaspalalaspasnarvalille 300
901 TCCGTTGCGTATGACGCGCGATTCACAAAGCGCGCGCATATTTT 950
|||||
301 SerglySerValleuasnlylalleThrhnglyalalehisasprlyle 317
951 GGGACGCTACCAACATCAGATTCGTTATCGAAGAGCGCGCGCAAG 1000
|||||
317 ugluArgylrhIsasnGlnlleSerValillegluLuglyargserlysg 334
1001 AGCTGTTCGGTGGTGGCGCGCGACGCGCAATPACCTCCATCGCGGT 1050
|||||
334 luLeuPhegllytrpValalalProglInProasplyslySerilleThrhArg 350
1051 ACGACCTCGCGCATTTCTGAAAAAACAACCTTCAGATTTCAGACAGC 1100
|||||
351 ThrThrleuGlyhisPheleuylsasnllyleuPheylsPheThrThal 367
1101 CGTCAACGCTGGCGACCGCGCATGTGGCGATTTGGTACTTACGAGCGG 1150
|||||
367 aValasnlyglylaspargalameValalProleeglyThryrgluArgv 384
1151 TATGCGCGTACAGACATTCGCTACCGCTTCGCGCATTAATGCTC 1200
|||||
384 alMetProleuasplleleuProThrleuLeuLeuargspheulleVal 400

```

1201 GCGATACGACAGCGCGAAGCATTTGGTTCGTAATGACGAGA.1250
 |||||
 401 GLYASPTInrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluG1 417
 1251 AAGCTCGCTTTGTCAGCTTGTCTGCGCGGCAATACGAATAGGCC 1300
 |||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434
 1301 CGCTGTGTGTAAGTGTGGAACATTTGAGAGAGAGCC 1341
 |||||
 434 roLeuLeuThrGlyValLeuGluThrIleGluLysGluGly 447
 seq_name: SwissProt_40:NORA_NEIMB

seq_documentation_block:
 ID NORA_NEIMB STANDARD; PRT; 447 AA.
 AC Q9K0M3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
 DE Na(+)-translocating NOR subunit A (Na(+)-NOR subunit A) (NOR complex subunit A) (NOR-1 subunit A).
 GN NORA OR NMB0569.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Hietelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Elsen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Neilson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parker D.S., Blatt E., Ciftone H., Clark E.B., Cotton M.D., Uitterlbeck T.R., Knouri H., Qin H., Vanathavan J., Gili J., Scariato V., Masignani V., Pizsa M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
 RT Science 287:1809-1815(2000).
 RL
 CC -|- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
 CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) + UBIQUINOL + NA(+)(OUT).
 CC -|- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORF, MORE AND MORE (BY SIMILARITY).
 CC -|- SIMILARITY: BELONGS TO THE NORA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE002412; AAF40997.1; -;
 DR TIGR: NMB0569; -;
 KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
 KW Complete proteome.
 SQ SEQUENCE 447 AA; 48635 MW; BE17AA78439BE477 CRC64;

alignment_scores:
 quality: 2177.00 Length: 447
 Ratio: 4.982 Gaps: 0
 Percent Similarity: 97.763 Percent Identity: 94.855

alignment_block:
 US-09-303-518D-127 x NORA_NEIMB ..
 Align seg 1/1 to: NORA_NEIMB from: 1 to: 447

1 ATGATTAATAATCAAAAAAGGTCTAAACCTCCATCGCGGCGAGACCGGA 50
 |||||
 1 MetLeuLysIleLysLysGlyLeuAsnLeuProIleAlaIleLysArgProG1 17
 51 GCAAGTCATTTTANGACGGCGCGCTCATTTACGGAAGTGGCTGCTGGCG 100
 |||||
 17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34
 101 AAGATATATGCGCGGTATGCGCCCTNGATGAAGTCAAGAGAGCGATGCC 150
 |||||
 34 IugLutYrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50
 151 GTCAAAAAAGGCCAAGTGTGTTGAAGCAAAAAGNATCCGCGCGCTGGT 200
 |||||
 51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValAla 67
 201 GTTTACCGCGCGCTTTCAGGCAAAATCGCGCCATTCATCCGCGCGGAAA 250
 |||||
 67 LpheThrAlaProAlaSerGlyLysIleAlaIleIleHisArgGlyGlu 84
 251 AGGGCGTACTTCAGTGGTGTGATTCGCGTGAAGGCAACGACGAATC 300
 |||||
 84 YsarGlyValLeuGlnSerValIleAlaValGluGluLysAsnArgLysIle 100
 301 GAGTTGGAACGCTACGCGCGCGCAAGCGTTGGCAAACTTAAGCGCGGANGA 350
 |||||
 101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluG1 117
 351 ANNNNGNNGCAATCTGATCCATCCGCTTGTGAGCTGCGCTGCGTANCC 400
 |||||
 117 ValArgArgAsnLeuIleGlnSerGlyLeuThrAlaLeuArgThrA 134
 401 GTCCGTTACAGCAAAATCCGCGCGTGTGATGCGAGCGCGTCCGCACTTC 450
 |||||
 134 rGProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150
 451 GTCAATGCGATGACACCAATCCGCTNGCGGAGAGACCTGTGGTGTGAT 500
 |||||
 151 ValAsnAlaMetAspThrAsnProLeuAlaIleAspProThrValIleI1 167
 501 CAAAGAGCGCGNGANGATTTACAGCAGANGTGTGCTGATTTAGCCGTT 550
 |||||
 167 eLysGluAlaIleAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184
 551 TGACCGAGGTAATAATTCATGTGTGTGAAGCAGCTGGCGGACAGCGCGG 600
 |||||
 184 eutThrGluArgLysIleHisValCysLysAlaIleAlaLysAlaAspAlaPro 200
 601 TCTGAATAATGCTGCCAACAATCGAACACATGAATTCGGCGGCCGCAATC 650
 |||||
 201 SerGluAsnAlaIleAsnIleGluThrHisGluPheGlyLysProHisPsr 217
 651 GGGCGGTTGAGTGGGACGACCATTCATTTCAATTGACGCGGTGCGCAA 700
 |||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234
 701 ACAAAACGTTTGACATCATTAATTCAGATGATTAATTCGCAATCGCGAGGT 750
 |||||
 234 snLysThrValTyrPheThrIleAsnTyrGlnAspValIleThrIleGlyArg 250
 751 TTTGTTGCAACAGCGCGTCTGAACACGAGCGGTGATGCTTTGGGTGG 800
 |||||
 251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyG1 267
 801 TTTCTCAAGTCACAAACCAACGCGCTTGGCGTACGCTTTGGGTGCGAAG 850
 |||||
 267 ySerGlnValaLysLysProAlaGlyLeuLeuArgThrValLeuGlyAlaLysV 284

```

851 TATCGCAATTACTGGGGGAATGGTTGACGACAGCAACCGCGATT 900
|||||
284 aISerGlnIleThrAlaGlyLeuValAspThrAspAsnArgValIle 300
|||||
901 TCCGGTTCGGTATTGAAGCGCGGATTACACAGCGCGCGCATTAATT 950
|||||
301 SerGlySerValIleuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrIle 317
|||||
951 GGGACGCTTACCACATGATGATTTCCGTTATCGAAGAGCGCGCAAG 1000
|||||
317 uGlyArgTyrHisAsnGlnIleSerValIleGlnGlnGlyArgSerLysG 334
|||||
1001 AGCGTTGGGCTGGGTGGCGCGCGCGCAAAATACCTCCACACGCGT 1050
|||||
334 IuLeuPheGlyTyrPValAlaIleProGlnProAspLysTyrSerIleThrArg 350
|||||
1051 ACGACCTCGGCGCATTTCTGAAAAACAACTCTCAAGTTACAGCAGC 1100
|||||
351 ThrThrLeuGlnHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367
|||||
1101 CGTCAACGGTGGCGACCGCGCATGGTGGCGGATTGTTACTTACAGCGCG 1150
|||||
367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGlnArgV 384
|||||
1151 TTAATCGCGTACAGACATCTGCTACCTGCTTTGGCGGATTAAAGCTC 1200
|||||
384 aIMetCProLeuAspIleuProThrIleuLeuAsnArgAspIleVal 400
|||||
1201 GCGCATACCGACAGCGCGCAAGCATTTGGTGTGTTGAATGAGCAAGA 1250
|||||
401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGlnLeuAspGlnG 417
|||||
1251 AGACCTGCTTTGTCAGCTTCGTCCTCCCGGCGCAATACGAATANGCC 1300
|||||
417 uAspLeuAlaLeuLysSerPheValCysProGlyLysTyrGlnTyrGly 434
|||||
1301 CGCTGTTGCGTAAGTCTGGAACNTTGAAGAAAGAAAGC 1341
|||||
434 rOLEuLeuArgLysValLeuGlnThrIleGlnLysGlnGly 447
|||||
seq_name: SwissProt_40:NORA_PASMU
seq_documentation_block:
ID NORA_PASMU STANDARD: PRT; 446 AA.
AC 09CUB1:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE (Na(+)-translocating NADH-quinone reductase subunit A (Ec 1.6.5.-)
DE (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
GN NORA OR PM1328
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PM70.
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC - FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC UBIQUINOL + NA(+)(OUT).
CC - SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE
CC AND NORF (BY SIMILARITY).

```

```

CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006171; AK03412.1;
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 446 AA; 48168 MW; 5C5287CB77C4932D CRC64;

alignment_scores:
Quality: 1614.50 Length: 447
Ratio: 4.087 Gaps: 1
Percent Similarity: 88.367 Percent Identity: 69.351

alignment_block:
US-09-303-518d-127 x NORA_PASMU
Align seg 1/1 to: NORA_PASMU from: 1 to: 446

1 ATGATTAAATCAAAAAAGTCTAAACCTGCCCATCGCGGACAGACGGA 50
|||||
1 MetIleThrIleLysGlyLeuAsnLeuProIleSerGlySerProG 17
51 GCAAGTCATTTATGACGGGCGCGCATTAACCAAGTCGCTTCTGGCG 100
|||||
17 uGlnValIleArgAspGlyAsnAlaIleThrGlnValAlaLeuLeuGlyG 34
101 AAGATATGCGCGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
|||||
34 IuGlnTyrValGlyIleLysArgProSerMetLysValArgGlnGlyAspVal 50
151 GTCAAAAAAGGCCAAGTGTGTTGAAGCAAAAAAGNATCCGGCGCGT 200
|||||
51 ValLysLysGlyGlnValLeuPheGlnAspLysAsnProGlnValVal 67
201 GTTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 250
|||||
67 IPherThrAlaProAlaSerGlyThrValThrAlaIleHisArgGlyAlaL 84
251 AGCGGTACTTCAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATC 300
|||||
84 yAsArgValLeuGlnSerValProAlaValAlaSerIleThrProValSerIlePhe 150
301 GAGTTCGAACGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 350
|||||
101 ThrPheGlnLysTyrThrThrGlnLeuAsnGlnLeuThrIleThrSerGlnG 117
351 ANTNNGNNGCATCGATCGATCCGCGTTGCGACTGCGCGCGGTANCC 400
|||||
117 nValArgGlnAsnLeuGlnAlaSerGlyLeuTyrPheThrAlaLeuArgThrA 134
401 GTCCGTTACAGAAAATCCCTGCCGTCGATGCCGCGCGCGCGCGATCTTC 450
|||||
134 rGProPheSerLysValProAlaValAlaSerIleThrProValSerIlePhe 150
451 GTCAATGCGATGACCAACCAATCCGCTGCGCGCGCGCGCGCGCGCG 500
|||||
501 CAAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 550
|||||
167 IeGlnIleSerAlaGlnAlaPheGlnAlaGlyLeuThrValLeuSerArgL 184
551 TGACCGAGCGTAATCAATGTTGTTGAAGCAGTGGCGCGCGCGCGCG 600
|||||
184 euHisGlnGlyLysValTyrLeuGlyLysAlaAlaAsnAlaSerIlePro 200

```

```

601  TCTGAAATGCTGCCACATGCAACACATGATTCGGCGCCGCGATCC 650
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
201  SerProserIleAlaIasnLeuAspValLysGluPheAlaGlyProHisIrr 217
651  GCGCGGTTGAGTGGCAGCGACATTCATTCATGAGCGCGTGGTSCAA 700
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
217  cAlaIlyLeuSerGlyThrHisIleHisPheLeuAspProValSerAlaT 234
701  ACAAAAGCGTTGGACATTCATTCATGATTCATTCATTCATTCATTCAT 750
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
234  hLysPheValIrrPyrIleAsnIrrGlnAspValIleAlaValGlyLys 250
751  TTGTTGGACAGCGCGCTGTGACACCGAGCGCGATTCGTTGGTGGGG 800
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
251  LeuPheThrThrGlyLysLeuAspValSerIrrValIalSerLeuAlaG 267
801  TTCTCAATGCAACAAACACCGCTTCGTTACCGGTTGGTGGGGAAG 850
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
267  yProGlnValLysAsnProAlrGlyLeuValAlrGlyValLeuGlyAla 284
851  TATCGCAATTCATTCGCGCGCAATTCGTTGACGAGACACCGCGTAT 900
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
284  euserGlnLeuThrAlaAsnGlnValLysAspGlyGlnAspAlrGlyVal 300
901  TCCGGTTGGTATTGACAGCGCGCATTCACAGAGCGCGCAGATATTT 950
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
301  SerGlySerValLeuSerGlyAlaLysAlaAlaGlyProValAspTyrLe 317
951  GGGAGCGTACCAACATGATTCGTTGATTCGAAAGAGCGCGCAAG 1000
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
317  uGlyAlrGlyAlaLeuGlnValSerValLeuGlnGlnGlyAlrGlyLys 334
1001  AGCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1050
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
334  LuPheLeuGlyTrrIleMetProGlyAlaAsnLysTyrSerLeuSerIrr 350
1051  ACGAGCGTGGCGCATTCCTGTAACAAACATTCATTCATTCATTCAT 1100
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
351  ThrValLeuGlnHisPheSerLys...LysLeuPheAsnPheThrIrrAl 366
1101  CGTCAACGCGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
366  AlaAsnGlnGlyGlnAlaGlnAlaMetValProIleGlyAlrGlyLys 383
1151  TAATGCCCTAGACATTCCTGCTACCGCTTCGTTGCGGATTCATGTC 1200
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
383  alMetProLeuAspIleIrrProThrIrrLeuLeuAlrGlyAspLeuAla 399
1201  GCGCATACCGAGCGCGCGCAAGCATTCGTTGATTCGAAATTCGAGAGA 1250
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
400  GlyAspThrAspSerAlaGlnAlaLeuGlyLysLeuGlnLeuAspGln 416
1251  AGACCTGCGCTTGGCAGCTTCGCTGCGCGCGCGCGCGCGCGCGCG 1300
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
416  uAspLeuAlaLeuGlyThrPheValLysProGlyLysAsnGlnLysGly 433
1301  CGCTGTTGCTAGAGTGGTGGAAACGNTTGAGAGGAGG 1341
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
433  rOleuLeuAlrGlnAlaLeuAspLysIleGlnLysGlnGly 446

seq_name: SwissProt_40:NORA_HAEIN
seq_documentation_block:
ID  NORA_HAEIN  STANDARD:  PRT:  447  AA.
AC  P43955: P43955:
DT  01-NOV-1995 (Rel. 32, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE  (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE  subunit A) (NOR-1 subunit A).
GN  NORA OR H10164.

```

```

OS  Haemophilus influenzae.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Haemophilus
OX  NCBI_TaxID=727;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RD / KW20 / ATCC 51907;
RX  MEDLINE=95350630; PubMed=7542800;
RA  Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA  Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA  Weidman J.F., Phillips C.A., Spriggs C.A., Hedblom E., Cotton M.D.,
RA  Utterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA  Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA  Venter J.C.;
RT  "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT  Rd."
RL  Science 269:496-512(1995).
RN  [2]
RP  REVISIONS.
RA  White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RL  Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE OF 1-75 FROM N.A.
RC  STRAIN=RD / KW20 / ATCC 51907;
RA  Barcak G.J., Helmer S.R.;
RL  Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN  [4]
RP  IDENTIFICATION AS NOR SYSTEM.
RC  STRAIN=RD / KW20 / ATCC 51907;
RX  MEDLINE=96176316; PubMed=8601449;
RA  Hayashi M., Nakayama Y., Unemoto T.;
RT  "Existence of Na+-translocating NADH-quinone reductase in Haemophilus
RT  influenzae."
RL  FEBS Lett. 381:174-176(1996).
RN  [5]
RP  IDENTIFICATION BY MASS SPECTROMETRY.
RX  MEDLINE=20137488; PubMed=10675023;
RA  Langen H., Takacs B., Evers S., Berndt P., Lahn H.W., Wipf B.,
RA  Gray C., Fountoulakis M.;
RT  "Two-dimensional map of the proteome of Haemophilus influenzae."
RL  Electrophoresis 21:411-429(2000).
CC  -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC  OF NA(+)-IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC  ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC  URISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC  UBIQUINOL + NA(+)(OUT).
CC  -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, MORE
CC  AND NORE (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC  -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC  FRAMESHIFT IN POSITION 63.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U32702; AAC21836.1; -
DR  EMBL: U20229; AAA62134.1; ALT_FRAME.
DR  TIGR: H10164; -
KW  Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW  Complete proteome.
SQ  SEQUENCE 447 AA; 48884 MW; 4670C93FB5FF0912 CRC64;

```

alignment_scores:

Quality: 1574.00 Length: 448
 Ratio: 4.057 Gaps: 2
 Percent Similarity: 86.607 Percent Identity: 67.634

alignment_block:

US-09-303-518D-127 x NORA_HAEIN ..

Align seg 1/1 to: NORA_HAEIN from: 1 to: 447

```

1 ATGATTAATAAATAAAGTCTTAACCTGCCATCGCGGCACAGCCGA .50
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleThrIleLysGlyLeuAspLeuProIleAlaGlyLysProAl 17
51 GCAAGTCATTATGACGGGCCCATTAACGAGTGGCTGGTGGCG 100
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 aglnvalIleHisSerGlyAsnAlaValAsnGlnValAlaIleLeuGly 34
101 AAGAATATGCGGATATGCGCCCTNGATGAAGTCACAGAGCGCATGCC 150
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 IuGlUtyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50
151 GTCAAAAAAGCCCAAGTCTGTGTAAGACAAAAAGNATCCGGCGTGGT 200
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValLysLysGlyGlnValLeuPheGlnAspLysAsnProGlyValAl 67
201 GTTTCACCGCGCGNGTTCAGGCAAAATGCGCCCATCGATCGCGCGCAA 250
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 ePheThrAlaProAlaSerGlyThrIleThrAlaIleAsnArgLysGlu 84
251 AGCGGCTACTTCAGTCGGTGTGATGCGCTGTAAGGCAACAGCAAAATC 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysArgValLeuGlnSerValValIleAsnValGluGlyAspGluLysIle 100
301 GAGTTCGAACGCTACGCCCGGACGCTGGCAAACTTAACGGCGGANGA 350
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ThrPheAlaLysTyrSerThrGlnIleLeuAsnThrLeuSerSerGlu 117
351 ANTNNNGNCAATCTGATCCAATCCGTTTGTGACTCGCTGCGTANCC 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValLysGlnAsnLeuIleGlnSerGlyLeuThrPheAlaLeuArgThr 134
401 GTCCGTTACAGCAAAATCCCTGCGTATGCCAGCGCTTGCCTTC 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rgProPheSerLysValProSerIleGlnSerGluAlaSerSerIlePhe 150
451 GTCAAGGGATGACCAATCCGCTNGCGAGACCTCGTGGTGTAT 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProSerValVal 167
501 CAAGAAGCGCGNGANGATTTCAGACGANGTNGCTGTATTCAGCCGTT 550
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 ulysGluTyrSerGlnAspPheThrAsnGlyLeuThrAlaLeuSerArg 184
551 TGACCGAGCGTA..ATCCATGTGTGTAAGCAGCTGGCGGACAGCTG 597
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euPheProSerLysProLeuHisLeuCysLysAlaGlyAspSerAsnIle 200
598 CGGTGAAATGCTGCCAATCGAACAATGAATGCGGCGGCCCA 647
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 ProThrAlaAspLeuGlnAsnLeuGlnIleHisAspPheThrGlyVal 217
648 TCCGCGCGGTTGAGTGGCAGCAGCATTCATTGATGCGCGTCCGTTG 697
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 sPheAlaGlyLeuValGlyThrHisIleHisPheIleAspProValAl 234
698 CAACAAAAACGTTTGGACCATCAATATATCAAGATGTAATGCCATGGA 747
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 IeGlnLysThrValThrHisIleAsnTyrGlnAspValAlIleAlaVal 250
748 CATTGTGTTGCAACAGCGCGTGTGAACACGCGCGGATTCGTTGGG 797
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 LysLeuPheThrThrGlyLeuLeuTyrSerGluArgValIleSerLeuAl 267

```

```

798 TGCTCTCAAGTCACAAACACGCTTGTGCGTACCGTTTGGGTGCGA 847
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 aglyProGlnValLysGluProArgLeuValArgThrThrIleGlyAla 284
848 AAGTATCGCAAAATTACTGCGGCGAATTGTTGACGACAGCAACCGCTG 897
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 snLeuSerGlnLeuThrGlnAsnGlnLeuSerAlaGlyLysAsnArgVal 300
898 ATTTCCGGTTCGGTATTAAGCGCGGATTAACAAAGCGCGCAGCATTA 947
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 IleSerGlySerValLeuCysGlyGlnIleAlaLysAspSerHisAsp 317
948 TTTGGACGCTACACACATTCAGATTTCGTTATCGAAGGCGCGACGA 997
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 rLeuGlyArgTyrAlaLeuGlnValSerValIleAlaGlnGlyAsnGlu 334
998 AAGAGCTGTTCCGCTGGTGGCGCGCGACCGGACAAATATCCATCAG 1047
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 ysgLysPhePheGlyTyrIleMetProGlnAlaAsnLysTyrSerVal 350
1048 CGTACGACCGCTGCGCCATTTCCTGAAAAACAAACTCTCAAGTTCAGC 1097
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ArgThrValLeuGlyHisPheSerLys...LysLeuPheAsnPheThr 366
1098 AGCGGTCAACGGTGGCGACCGCGCATGGTGCATGGTGTACTTACGAG 1147
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
366 rSerGluAsnGlyGlyGluArgAlaMetValProIleGlySerTyrGlu 383
1148 GCGTAATCCGCTACAGACATCTGCTACCGCTGTTGGCGCATTTAATC 1197
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
383 rGlyAlaMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeu 399
1198 GTCGGCGATPCCGACGACGCGCAAGCATTGGTGGTCTGGAATTGAGCA 1247
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 ValGlyAspThrAspGlyAlaGlnGlnLeuGlyCysLeuGlnLeuAsp 416
1248 AGAAGACTTCGCTTGTGACGCTTGTGCGCGCGCAATATCGAATTANG 1297
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 uGlnAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyr 433
1298 GCCCGCTGTGCGTAGGCTGGAACACNTGAGAGGAAGGAC 1341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 LysIleLeuArgGlnValLeuAspLysIleGluLysGlu 447
seq_name: Swissprot_40:NORA_VIBHA
seq_documentation_block:
ID NORA_VIBHA STANDARD; PRT; 446 AA.
AC Q9RFW1;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+))-translocating NQR subunit A (Na(+))-NQR subunit A (NQR complex
  subunit A) (NQR-1 subunit A).
OS NORA.
GN Vibrio harveyi.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB120;
RX MEDLINE=20056044; PubMed=10587447;
RA Zhou W., Bertsova Y.V., Feng B., Tsatsos P., Verkhovskaya M.L.,
  Gennis R.B., Bogachev A.V., Barchera B.;
RT "Sequencing and preliminary characterization of the Na+-translocating
  NADH:ubiquinone oxidoreductase from Vibrio harveyi.";
RL Biochemistry 38:16246-16252(1999).
CC -!- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
  UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
  OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQRE
  ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF

```

CC UBISEMIQUINONE TO UBIOQUINOL.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIOQUINONE + NA(+)(IN) = NAD(+) +
 CC UBIOQUINOL + NA(+)(OUT).
 CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
 CC AND NQRF (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NQRA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF165980; AAF15411.1;
 CC DR Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport.
 CC KM
 CC SEQUENCE 446 AA; 48365 MW; 98058388BID7A97C CRC64;

alignment_scores:

Quality: 1424.50 Length: 447
 Ratio: 3.759 Gaps: 1
 Percent Similarity: 84.787 Percent Identity: 61.745

alignment_block:

US-09-303-518D-127 x NQRA_VIBHA ..

Align seg 1/1 to: NQRA_VIBHA from: 1 to: 446

1 AAGATAAATAAATAAAGGTTAAACCTGCCATCGCGGAGACCGGA 50
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1 MetIleThrlleLysLysGlyLeuAlaLeuProIleAlaGlyAlaProse 17
 51 GCAAGCATTTATGACGGCGCGTCAATACCGAAGTCCGCTGGTGGC 100
 :|||: ||||| ||||| :|||: :|||: :|||: :|||: :|||:
 17 rclnValIleasnAspGlyLysThrIleLysLysValAlaLeuLeuGly 34
 34 LGLIutylValGlyMetArgProIleMetHisValArgValGlyLys 50
 101 AAGAAATATGCGGTATGCGCCCTTGATGAAGTCAAGAGAGCGATG 150
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 34 LGLIutylValGlyMetArgProIleMetHisValArgValGlyLys 50
 151 GTCAAAAAAGCCAAAGTGTGTTGAAGCAAAAAGNATCCGGCGGTGT 200
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 51 ValLysLysAlaGlnValLeuPheGlnLysAsnProGlyValLys 67
 201 GTTTACCGCGCCGTTTACGCAAAATGCGCCCATCCATCGCGCGCAA 250
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 67 spherThrAlaProAlaAlaGlyLysValIleGlnLeuAsnArgGlyAla 84
 251 AGCGGCTACTTCAGTCGGTGTGATGCGGTGAAGGCAAGCAAGATC 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 84 ysaArgValLeuGlnSerValValIleGlnValAlaGlnGlyGlnGlnVal 100
 301 GAGTTCGAACGCTAGCGCCCGCAAGCGTTGGCAAACTTAAGCGCGG 350
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 101 ThrPheAspLysPheGlnAlaAlaGlnLeuAlaGlyLeuAsnArgGly 117
 351 ANTNNNGNCAATGATCAATCCGTTTGGACCTGCGTGGCGTGAAGC 400
 :|||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
 117 IleLysThrGlnLeuValGlnSerGlyLeuThrPheAlaLeuAlaGln 134
 401 GTCCGTTAGCAAAATCCCTCGCGTGCATGCGGAGCGGTCGCATTC 450
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 134 rgrProPheSerLysValProAlaIleGlnSerAlaThrLysAlaIlePhe 150
 451 GTCAATGGATGAGACCAATCCGCTGCGGACACACCTGTGGTGGAT 500
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 151 ValIleAlaLeuAspThrAsnProLeuAlaAlaGlnProGlnLeuIle 167
 501 CAAGAGACCCGNCGANGATTTGACAGCAGANGTNTGCTGATTGAGCG 550
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 167 ethrGlnGlnGlnGlnAlaPheValAlaGlyLeuAspIleLeuSerAla 184

551 TGACCGAGCGTAAATCCATGCTGTANGAGCAGCTGGCGGAGACGTCCG 600
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 184 eutrhGlnGlyLysValIleValCysLys...SerGlyThrSerLeuPro 199
 601 TCTGAAAATGCTGCCACATCGCAAAACATGATTCGGCGCGCCGATCC 650
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 200 SerSerSerGlnSerAsnValGlnGlnHisValPheAspGlyProHisPr 216
 651 GCGCGCTTGAGTGGCAGCAGCATTCATTCATTGAGCGCGTGGTCAA 700
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 216 OAlaGlyLeuAlaGlyThrHisMetHisPheLeuTyProValAsnAlaG 233
 701 ACAAACCGTTGGACCATCAATTAATCAAGTGAATTCATTCGATCGAC 750
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 233 IuasValAlaIleArgSerIleAsnTyGlnAspValIleAlaPheGlyL 249
 751 TTTGTTGCAACAGCCCGTGTGACACCGAGCGCGTGAATGCTTGGGTG 800
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 250 LeuPheLeuThrGlyGlnLeuTyThrAspArgValValSerLeuAlaG 266
 801 TTTCAAGTCAACAACACCGCGCTTGGTACCGTTTGGGTGCGAAG 850
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 266 yProValValAsnAsnProArgLeuLeuArgThrValValGlyAlaSer 283
 851 TATCGCAATTCCTCGCGGCGAATGGTTGACGCGACACACCGCGTAT 900
 :|||: ||||| ||||| :|||: :|||: :|||: :|||: :|||:
 283 eugLiasPheThrAspSerIleLeuMetProGlyGlnValArgValIle 299
 901 TCCGTTTCGATTAATGAACGGCGCGATTAACAAGCGCGCGATTAAT 950
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 300 SerGlySerValLeuSerGlyThrGlnAlaSerGlyProHisAlaTyLe 316
 951 GGGACGCTACCAATCAATATTCGTTATGAGAAGAGCGCGAGCAAG 1000
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 316 uGlyArgTyHisGlnGlnValSerValLeuArgGlyGlyArgAspLys 333
 1001 AGCTGTTCGCGTGGTGGCGCGCGGACGCAATCTCCATCAGCGCT 1050
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 333 IuleuPheGlyTTPAlaThrProGlyLysAsnLysPheSerIleThrLys 349
 1051 ACGACCTCGCGCATTTCTGAAAAACAACATCTCAAGTTCAGCAGC 1100
 :||: ||||| ||||| :||: :||: :||: :||: :||: :||:
 350 SerPheLeuGlnHisIlePheLysGlyGlnLeuPheAsnMetThrThr 366
 1101 CGTCAACGGTGGCGACCGCCCATGCGTGGATTTGATTCAGACGCG 1150
 :|||: ||||| ||||| :|||: :|||: :|||: :|||: :|||:
 366 rThrAsnGlySerAspArgAlaMetValProIleGlyAsnTyGlyArg 383
 1151 TTAATGCCGTAGACATCTGCGTACCCCTGCTTTGGCGGATTAATC 1200
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 383 alMetProLeuAspMetGlnProThrLeuLeuLeuArgAspLeuCyAla 399
 1201 GCGGATACCGACAGCGCGCAAGCATTTGGTGTGAATTTGAGCAAGA 1250
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 400 GlyAspThrAspSerAlaGlnThrLeuGlyAlaLeuGlnLeuAsnGln 416
 1251 AGACCTTCGCTTTGAGAGCTTCTGCTGCGCGGCAATACGAATANG 1300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 416 uAspLeuAlaLeuLysThrPheValLysProGlyLysTyGlyTyGlyG 433
 1301 CGCTGTTCGCTAAGTGTGGAACACNTTGAGAAGAAGAC 1341
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 433 IueLeuLeuArgGlyLysLeuAspThrIleValLysGlnGly 446

seq_name: SwissProt_40: NQRA_VIBHA

seq_documentation_block:

ID NQRA_VIBHA STANDARD; PRT; 446 AA.
 AC 05686;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)


```

601 TGTGAAATGCTCCACATCGAAGACATGATTCGGCGCCGACATCC 650
      :::::|||||  |||  |||  |||  |||  |||  |||
200 ArgSerSerGlnSerAsnValGluGlnValPheAspGlyProHisSpr 216
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
651 GGGCGGTTGGTGGCAGCAGCATTCATTCATTCAGCCGCGTGGTGA 700
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
216 GAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 233
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
701 ACAAACGGTTGGCAGCATTCATTCATTCATTCATTCATTCATTC 750
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
223 TAAATGATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 249
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
751 TTTGTTGCAACAGCGCGTGTGAACAGCGCGGTGATTCGTTGGTGG 800
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
250 LeuPheLeuThrGlyLeuLeuThrAspArgValValSerLeuAlaG 266
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
801 TTTGCAAGTCAACAGCGCGTGTGAACAGCGCGGTGATTCGTTGGT 850
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
266 YProValValAsnProAlaGlyLeuValArgThrValIleGlyAlaSer 283
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
851 TATCGCAATTCATTCGCGGCGATTCGTTGACGAGACAGCGCGGTG 900
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
283 euAspAspLeuThrAspAsnGluLeuMetProGlyGluValArgValIle 299
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
901 TCCGTTGCGTATTCGAACGCGCGATTCACAGCGCGCGCATTTAT 950
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
300 SerGlySerValLeuThrGlyThrIleAlaThrGlyProHisAlaTyr 316
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
951 GGGAGCGTACAGCATTCATTCGTTGTTGGAAGAGCGCGCGCAAG 1000
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
316 uGlyArgGlyHisGlnGlnValSerValLeuArgGlyArgGlyGly 333
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1001 AGCTGTTGCGGTTGCGCGCGCGCGCGCGCAATTCATTCATTC 1050
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
333 IuLeuPheGlyTyrAlaMetProGlyLysAsnLysPheSerValThrArg 349
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1051 ACGAGCGTGGCGCATTCCTCGAAGAAACAACTCTCAAGTTACAG 1100
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
350 SerPheLeuGlyHisValPheLysGlyGlnLeuPheAsnMetThrThr 366
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1101 CGTCAACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
366 rThrAsnGlySerAspArgSerMetValProIleGlyAsnTyrGluArg 383
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1151 TAATCGCGTGAAGCATTCCTGCTACCTGCTGTTGCGCGATTAAT 1200
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
383 alMetProLeuAspMetGluProThrLeuLeuArgAspLeuGlyAla 399
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1201 GGGCGTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1250
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
400 GlysAspThrAspSerAlaGlnAlaLeuGlyAlaLeuGlnLeuAspGlu 416
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1251 AGACCTGCTTGTGAGCTTGTGCTGCTGCGCGCGCAATTAAGT 1300
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
416 uAspLeuAlaLeuGlyThrPheValGlyProGlyLysTyrGluTyrGly 433
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
1301 CGCTGTTGCGTGAAGTCTGGAACCTTGAAGAGAGAGCG 1341
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
433 hrLeuLeuArgGlyLysLeuAspThrIleGlyLysGly 446
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
seq_name: swissprot_40:NORA_VIBCH
seq_documentation_block:
ID NORA_VIBCH STANDARD: PRT: 446 AA.
AC Q9KPS1; Q9X4Q3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE subunit A) (NOR-1 subunit A) (NOR subunit A) (NOR complex
GN NORA OR VC2295.
Vibrio cholerae.

```

```

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RX NCBI_TaxId=666.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=99179036; PubMed=10077658;
RA Haese C.C., Mekalanos J.J.;
RT *Effects of changes in membrane sodium flux on virulence gene
expression in Vibrio cholerae.
Proc. Natl. Acad. Sci. U.S.A. 96:3183-3187(1999).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberger J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RT Nature 406:477-483(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
OF NA(+)- IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NORE
ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
URISEMIQUINONE TO UBIQUINOL.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
UBIQUINOL + NA(+)(OUT).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORD, NORE
AND NORF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF117331; AAD29962.1;
DR EMBL: AE004300; AAP95439.1; ALT_INT.
DR TIGR: VC2295;
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
Complete proteome.
SO SEQUENCE 446 AA; 48624 MW; 428E8C397EBA163D CRC64;

```

alignment_scores:

Quality: 1412.50 Length: 447
 Ratio: 3.767 Gaps: 1
 Percent Similarity: 83.893 Percent Identity: 60.403

alignment_block:

US-09-303-518d-127 x NORA_VIBCH ..

Align seg 1/1 to: NORA_VIBCH from: 1 to: 446

```

1 ATGATTAATCAAAAAAGGTCTAATACCTGCCGCGGCGAGACGGA 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||
1 MetIleThrIleLysGlyLysLeuAspLeuProIleAlaGlyThrProse 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||
51 GCAAGTCATTTATGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 100
|||||  |||||  |||||  |||||  |||||  |||||  |||||
17 rGlnValIleSerAspGlyLysAlaIleLysValAlaLeuLeuGly 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||
101 AAGATTAAGCGCGTATGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
|||||  |||||  |||||  |||||  |||||  |||||  |||||
34 LuGlyLysValGlyMetArgProThrMetHisValArgValGlyAspGlu 50

```

```

151 GTCAAAAAGCCAGTGTCTTTGAAGACAAAAGNATCCGGCGTGT 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 VallysYsAlaGlnIleLeuPheGluAspIlybLysAsnProGlyVally 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 GTTACCGCGCCNGTTTCAGCAAAATGCCGCCATCCATCGCGCGAAA 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 sPheThSerProValSerGlyLysValValGluIleAsnArgGlyAla 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 AGCCGCTACTCAGTGTGTGTATCCGTTGAGGACAGCAAGCAATC 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 YsATGValLeuGlnSerValValIleGluValAlaGlyAspAspGlnVal 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 GAGTTCAGACGCTACGCCGCCGAAGCCTTGCAAACTTAAGCGCGANGA 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 ThrPheSpsYrPheGluAlaAsnGlnLeuAlaSerLeuAsnArgAspAl 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 ANINNGNCAATGTATCCATCCATCCGTTGTGACTGCGCTGCGTANCC 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 AlLeYsThrGlnLeuValGlnSerGlyLeuTrpThrAlaPheArgThra 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 GTCCGTACGCAAAATCCCTGCGTGCATGCGCGCTGCGCATTC 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 rGpPheSerLysValProAlaIleAspSerThrSerGluAlaIlePhe 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 GTCAATGCGATGACACCAATCCGCTNGCGGACAGCCCTGTGTGTAT 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 ValThrAlaMetAspThrAsnProLeuAlaIleGluProThrValVal 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 CAAGAAGACCGNCANGATTTTCAGACAGANGTNGCTGTATGACCGGTT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 eAsnGlnGlnSerGluAlaPheValAlaGlyLeuAspValLeuSerAla 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 TGACGAGCGTAAATCCATGCTGTGAAGACAGTGCAGAGAGTCCG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 eunThrGlyLysValIlyValIlyLysLys..LysGlyThrSerLeuPro 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 TCTGAATATGCTGCCACATCGAAACATGAATTCGCGCGCCGATCC 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 ArgSerGlnGlnProAsnValGlnGlnIleValPheAspGlyProHisPr 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 GCGCGGTTGAGTGGACGACGACATTCATTTATGAGCCGCTGCGTGCA 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 cAlaGlyLeuAlaGlyThrHisMetHisPheLeuTrpProValSerAla 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 ACAAAACGCTTGACCATCAATATCAAGATGAATTCATCCATGGAGCGT 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 sPhsValAlaIleTrpSerIleAsnTrpGlnAspValIleAlaValGlyGln 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 TTGTTTGCAACAGCGCGCTGTGAACAGCGCGCTGATTCGTTGGGTG 800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 LeuPheLeuThrGlyGluLeuTrpThrGlnArgValValSerLeuAlaG 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 TTCTCAAGTCAACAAACAGCGCTCTGCGTACCGCTTTGGTCCGAAG 850
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 yProValValAsnLysProArgLeuValArgThrValMetIlyAlaSer 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 TATCGCAATTAATCGCGGCAATTTGTTGACGACAGACACCGCGTAT 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 eugInGlnLeuValAspSerGluIleMetProGlyGluValArgIleIle 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
901 TCCGGTTCGATTTGACGCGCGATTAACAGAGCGCGCGCATATTT 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
300 SerGlySerValLeuSerGlyThrLysAlaThrGlyProHisAlaIly 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
951 GGGAGCGTACACATCAGATTCGTTATGAGAGAGCGCGCGCAAG 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
316 uGlyArgThrHisLeuGlnValSerValLeuArgGlnIlyArgAspLys 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1001 AGCTTTGGCGTGGTGGCGGACGCGCAATAATCCATCCATCGCGGT 1050
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
333 lueuPheGlyTrpAlaMetProGlyLysAsnLysPheSerValThrArg 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1051 ACGACCTCGCGCATTTCTGAAAAACAACCTTCAAGTTCCAGACAGC 1100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

350 SerPheLeuGlyHisLeuPheLysGlyGlnValIlyAsnMetThrThr 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1101 CGTCAAGGTGGCGCGCGCCGATGTCGCGATTCGCTTACGAGCGC 1150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
366 rThrsnGlySerAspArgSerMetValProIleGlyAsnTrpGlyLysV 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1151 TAATGCCGCTAGACATCCCTACCCGCTTTTGCGGATTAATGCTC 1200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
383 AlMetProLeuAspMetGluProThrLeuLeuArgAspLeuLysAla 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1201 GCGCATACCGACAGCGCGCAGCATTTGGTGTCTTGGAATGAGCAAG 1250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
400 GlyAspSerAspSerAlaValArgLeuGlnIlyAlaLeuGlnLeuAsp 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1251 AGACCTCGCTTGTGACGCTGCTGCGCGCGCAATATGCAATANGSC 1300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
416 uAspLeuAlaLeuCysThrPheValCysProGlyLysTrpGlyGlyG 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1301 CGCTGTTCGTAGAGTGTGGAACCTTGAGAGAGAGC 1341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 lueuLeuArgGlyLysLeuAspLysIleGlyLysGlyGly 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

seq_name: SwissProt.40:NORA_PSEAE

seq_documentation_block:
ID NORA_PSEAE STANDARD; PRT; 445 AA.
AC O9H2K6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
DE (Na(+)-translocating NOR subunit A) (Na(+)-NOR subunit A) (NOR complex
DE subunit A) (NOR-1 subunit A).
DE NORA OR PA2999.
GN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Lim Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Luan R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO MORE
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIQUINONE TO UBIQUINOL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+) (IN) = NAD(+) +
CC UBIQUINOL + NA(+) (OUT).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NORA, NORB, NORC, NORD, NORE
CC AND NORF (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF004724; AAC06387.1; -
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW

```

KW Complete proteome.
SQ SEQUENCE 445 AA; 48081 MW; ED0468D06559B701 CRC64;

alignment_scores:
Quality: 1298.00 Length: 447
Ratio: 3.576 Gaps: 1
Percent Similarity: 81.208 Percent Identity: 57.271

alignment_block:
US-09-303-518D-127 x NORA_PSEAE ..
Align seg 1/1 to: NORA_PSEAE from: 1 to: 445

```

1 ATGATTTAAATCAAAAAAGGCTTAACCTGCGCCATCGGGGCGACGACGCGGA 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleIysIleIleIysArgIleIleuAspIleuSerProIleSerGIyAlaIleProGI 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGIlnArGIleGIuAlaIlaIlaArgProValArgSerValAlaIleuIleGIyP 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 AGAATATATCGGGTATGGCCCTTGATGAAGAATCAAGTCAAGAGCGCATCC 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 heAspIyThISGIyMetIysProThrMetAlaValGIlnValGIyAspArg 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 GTCAAAAAAGCGCAAGTCTGTGTGAAGACAAAAAGNATCCGGGGGTGCT 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValIysIleuGIyGIlnValIleuPheThrAspIysIleuAsnProSerValSe 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCGCATCCATCGCGCGGAAA 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 IyThrIlaIleProGIyAlaGIyValIleSerAlaIleIleIleArgGIyIuI 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 AGCGGCTACTTCAGTCGTCGATTCGCGTGAAGGCAAGCAGACGAATC 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysArgValIleuGIlnSerValIleuAspIleuAspGIyAspGIuGIlnIleu 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GAGTTGACAGCCTACGCGCCGACGCGTGAAGCAAACTTAAGCGCGGANGA 350
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 GIuPheAlaIleArgIyProAlaIleAspIysIleuAlaIleuSerAlaIleGIuI 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 ANTNGNGNGCAATCGATCCATCCATCCGTTTGGACTGCGCTGCGGTANCC 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 nValaIleArgAspAsnIleuGIlnSerGIyIleuIyPThrAlaIleuArgIyThra 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 GTCCGTTAGCAAAATCCCTGCGCGCATCGCGGACGCGTTCGCATCTC 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 rGIProPheSerIyValIleProAspProGIuSerProSerIlePhe 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
451 GTCAATGCGATGACACCAATCCGTCNGCGCAGACCCCTGTGTTGTGAT 500
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ValIleThraIleAspThrGIlnProIleuAlaIleAspProGIlnValIleI 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 CAAGAAGACCCGNCANGATTTCAACACGANGTCTGTGATTCAGCCGTT 550
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eAlaGIuGIlnGIyGIuAlaIlePheGIlnAlaGIyIleuThValIleuGIyArgI 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 TGACCGAGCGTAAATCATGTGTGAAGCAGCTGGGCGGACGAGTCCGCG 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 euAla.....ArgValIleuGIlnSerGIyAspAlaGIuGIyValSerIleuPro 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGCGCGCATCC 650
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GIyGIuAlaIleuSerGIyValIleThraIleAlaIlePheSerGIyProIleSPr 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 GCGCGGTTTGAGTGCGACGACATTCATTCATTCATTCATTCATTCATTCATTC 700
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 oAlaGIyIleuProGIyThrIleIleIleIleIleIleIleIleIleIleIleIle 232
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 ACAAAACGTTTGACCATCAATATCAAGATCAATTCATTCATTCATTCATTCATTC 750
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 IyIysSerValIlePheAsnIleuAsnIyIleIleIleIleIleIleIleIleIle 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

751 TTGTTTGCACACAGCCCTGTGAACACCGGACCGCGGTATTCCTTGGGTGG 800
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 IeuPheThrIleGIlnGIlnIleuIyPThrGIlnArgValIleAlaIleuAlaGI 265
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
801 TTCTCAAGTCAACAAACCGCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 yProValIleGIlnIyIysProAlleuValArgIleArgIleuGIyAlaAsnI 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 TATCGCAAAATTAATCTCGCGCGCAATTCGTTGACCGACGACGACGCGTAT 900
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 euAspGIleuAlaIleGIlnGIlnIleuGIlnProGIyAsnAlaIleuIle 298
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
901 TCCGCTTCGCTATTCAGACGCGCGCATTCACAAGAGCGCGCATTCATTCAT 950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 SerGIySerValIleuGIlnGIyArgIleAlaIleIleGIyAlaIleArgIle 315
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
951 GGGACGCTTACCAATGATTCCTGTTTCAAGAAAGAGCGCGGACGACGAAAG 1000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 uGIyArgIyThIleIleuGIlnIleuSerCysIleuGIyAspGIlnArg 332
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1001 AGCTGTTGCGCTGGGTTGCGCGCAGCGGACGACGCAAAATCTCCATCAGCG 1050
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
332 IuPheIleuIleIyIleuArgAlaGIyValGIlnIyIleIleSerIleuAsn 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1051 ACAGACCTCGCGCATTCCTCGAAAAACAACTTCACAAGTTCACGACAC 1100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 ValPheValIleSerAlleuIleuGIlnGIyIysArgPheAlaIlePheThrI 365
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1101 CGTCAACGAGTGGCGACCGCGCCATGTCGCGGATTCGATTCAGACGCGG 1150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 rIleAsnGIySerProAlleuAlaMetValProValGIyAsnIyGIlnAlaI 382
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1151 TAAATCGCGTACAGATCTGCTGCTTACCTGCTGCTGCTGCTGCTGCTGCTG 1200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
382 AlMetProIleuAspIleuIleuProThrGIlnIleuIleuArgIyIleuIle 398
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1201 GGGGATACGACAGCGCGGACGATTCGGGCTGCTGGAATTCAGACGAA 1250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 GIyAspThrGIlnMetAlaGIlnIyIleuGIyAlaIleuGIlnIleuAspGI 415
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1251 AGACCTCGCTTGTGCGAGCTGCTGCTGCGCGGCAATTCGAATGAGCGC 1300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 uAspIleuAlaIleuGIlnSerGIyValCysAlaGIyIyIyGIlnIyGIyP 432
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1301 CGCTGTTGCGTAAAGTGTGGAACCTTGACAGAGGAGGC 1341
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
432 roIleIleuArgAspAsnIleuAlaArgIleGIlnGIlnGIy 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: SwissProt_40:NORA_CHLIPN

seq_documentation_block:

ID NORA_CHLIPN STANDARD: PRT: 467 AA.

AC 092762:

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable Na(+)-translocating NADH-quinone reductase subunit A

DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)

GN (NQR complex subunit A) (NQR-1 subunit A).

GN NORA OR NQR1 OR CPN0743 OR CPN002.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.

OX NCBI_TaxID=83556;

RA

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,

RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";

RL Nat. Genet. 21:385-389(1999).

[2]


```

1044 CAGCGTAGACCGCTGGCCATTCTGTAACAACTCTTCAAGTTCA 1093
      :|||||  |||  |||  |||  |||  |||  |||  |||  |||
358 efrhlysthrlyleuserglyphehlyslsystarg...thrtyt 374
      ||  |||  |||  |||  |||  |||  |||  |||  |||
1094 CG.....ACAGCGTCAGCGGCGGCGCCAGCGCCAGTGTGGCGGANT 1134
      ||  |||  |||  |||  |||  |||  |||  |||  |||
374 hrasnproasphrasnleuhsiglyglutthargproillelespthr 390
      ||  |||  |||  |||  |||  |||  |||  |||  |||
1135 GGTACTTACGAGCGCGTAATGCGCTAGACATCGCTCCCTGCTCTTT 1184
      |||  |||  |||  |||  |||  |||  |||  |||  |||
391 AsplletyAspIysValmetProleuargyleProvalProleu 407
      |||  |||  |||  |||  |||  |||  |||  |||  |||
1185 GGGCGATTATCGTCGCGGATACGACGCGCCGACGATTGGTGGT 1234
      |||  |||  |||  |||  |||  |||  |||  |||  |||
407 elysalavalilethlyasnphaspheualaasnleuglyphel 424
      |||  |||  |||  |||  |||  |||  |||  |||  |||
1235 TGGATTGGACGAGACGACCTGCTTGTGCGCTGCTGCTGCGCGGCG 1284
      |||  |||  |||  |||  |||  |||  |||  |||  |||
424 eugllyvalysglylualasphealaleuProthleuileasproser 440
      |||  |||  |||  |||  |||  |||  |||  |||  |||
1285 AATATGCAATANGCGCGCTGTGCTGTAAGTGTGGAACCTTGAGAA 1334
      |||  |||  |||  |||  |||  |||  |||  |||  |||
441 LysThrGluMetLeuThrIleValysGluSerLeuIleGluTyrAla 457
      |||  |||  |||  |||  |||  |||  |||  |||  |||
1335 GGAA 1338
      |||  |||  |||  |||  |||  |||  |||  |||  |||
457 SGLu 458
      |||  |||  |||  |||  |||  |||  |||  |||  |||

seq_name: SwissProt_40:NORA_CHLMU

seq_documentation_block:
ID NORA.CHLMU STANDARD; PRT; 465 AA.
AC OPRU3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable Na(+)-translocating NADH-guione reductase subunit A
DE (EC 1.6.5.-) (Na(+)-translocating NQR subunit A) (Na(+)-NQR subunit A)
DE (NQR complex subunit A) (NQR-1 subunit A).
GN NORA OR TC0002.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gysin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: NQR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. NORA TO NQR
CC ARE PROBABLY INVOLVED IN THE SECOND STEP, THE CONVERSION OF
CC UBISEMIOUONE TO UBIQUINOL (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE + NA(+)(IN) = NAD(+) +
CC UBIQUINOL + NA(+)(OUT).
CC -1- SUBUNIT: COMPOSED OF SIX SUBUNITS: NORA, NORB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NORA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).

```

```

CC EMBL: AE002269; AAF38896.1; -
DR TIGR: TC0002; -
KW Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
KW Complete proteome.
SQ SEQUENCE 465 AA; 51588 MW; 12A1AB6FA64F2C CRC64;

```

```

alignment_scores:
      Quality: 441.00      Length: 464
      Ratio: 1.521      Gaps: 14
Percent Similarity: 62.500      Percent Identity: 28.448

```

alignment_block:

US-09-303-518D-127 x NORA_CHLMU ..

Align seg 1/1 to: NORA.CHLMU from: 1 to: 465

```

4 ATTAATAATCAAAAAAGGTCTTAACCTGCGCGGCGGACGAGCGAGCA 53
  |||  |||  |||  |||  |||  |||  |||  |||  |||
3 IlelleValSerArglyLeuAspLeuSerLeuLysGlyLalProLysGI 19
  |||  |||  |||  |||  |||  |||  |||  |||  |||
54 AGTATTATGACGGG.....CCGTCATTACCGAAGTGGCGTGC 94
  |||  |||  |||  |||  |||  |||  |||  |||  |||
19 UserGlyPheCysGlyLysValAspProAlaPheValSerValAspLeu 35
  |||  |||  |||  |||  |||  |||  |||  |||  |||
95 TTGCGCAAGAAATATGCGCGTATGCGCGCCCTGATGAAGTCAAGAAAGC 144
  |||  |||  |||  |||  |||  |||  |||  |||  |||
36 ....ArgProPheAlaProleuProleuGlyValLysValSerProGly 50
  |||  |||  |||  |||  |||  |||  |||  |||  |||
145 GATCGCGTCAAAAAAGCGCAAGTGTCTGTAAGACAAAAAGNATCGGG 194
  |||  |||  |||  |||  |||  |||  |||  |||  |||
51 AsplInleThrIleArglySerProleuAlaGluTyrLysSerPheProGI 67
  |||  |||  |||  |||  |||  |||  |||  |||  |||
195 CGTGCTTTACCGCGCGCGCTTTCAGGCAAAATCGCGCATCATCGCG 244
  |||  |||  |||  |||  |||  |||  |||  |||  |||
67 YValPheIleThrSerSerValAspGlyGluValIleGluIleArgArg 84
  |||  |||  |||  |||  |||  |||  |||  |||  |||
245 GCGAAAGCGCGTACTTCAGTCCGTCGTGATGTCCTGTA...GGCAAC 291
  |||  |||  |||  |||  |||  |||  |||  |||  |||
84 LyserylSerArgAlaLeuLeuAspIleValIleLysLysLysProGlyVal 100
  |||  |||  |||  |||  |||  |||  |||  |||  |||
292 GACGAAATCGAGTTGCAAGCTACGCGCGCGCAAGCGTTGGCAACTTAA 341
  |||  |||  |||  |||  |||  |||  |||  |||  |||
101 SerGlnThrLysPheSer.....TyrAspLeuHisAlaLeu 113
  |||  |||  |||  |||  |||  |||  |||  |||  |||
342 CGGCGAAGAAATNNNGNCAATGATCAATCCAGTGTTCGACGCGC 391
  |||  |||  |||  |||  |||  |||  |||  |||  |||
113 rGlnLysGluLeuLeuGluValPheLysLysGluGlyLeuPheThrLeu 130
  |||  |||  |||  |||  |||  |||  |||  |||  |||
392 TGGGTATNCCTCGCTCAGCAAAATCCCTGCGTCGATGCGGAG...CCG 438
  |||  |||  |||  |||  |||  |||  |||  |||  |||
130 heLysGlnArgProPheAsn...IleProAlaLeuProthIrgInserPro 145
  |||  |||  |||  |||  |||  |||  |||  |||  |||
439 TTGGCATCTTCTGCAATGCGATGACACACCAATCGCTN.....GCGGC 482
  |||  |||  |||  |||  |||  |||  |||  |||  |||
146 ArgAspValPheIleAsnLeuAlaAspAsnArgProPheThrProserVa 162
  |||  |||  |||  |||  |||  |||  |||  |||  |||
483 AGACCCGTGTGTGTATCAAAAGACCCGCGCANAT.....T 520
  |||  |||  |||  |||  |||  |||  |||  |||  |||
162 IGIuLysHisLeuSerLeuPheSerLysGluAspGlyTyrTyrIleP 179
  |||  |||  |||  |||  |||  |||  |||  |||  |||
521 TCAGACGANGTNTGCTGATATGAGCGGTTCAGCGGTAATAAATCAT 570
  |||  |||  |||  |||  |||  |||  |||  |||  |||
179 heValValGlyValGlnAlaIleAlaLysLeuPheLysLeuProHis 195
  |||  |||  |||  |||  |||  |||  |||  |||  |||
571 GTGTGTAAGGCACTGCGCGACGAGTGCCTGGAATGTGCGCAACAT 620
  |||  |||  |||  |||  |||  |||  |||  |||  |||
196 IleValSerThrAspArgLeuSerLeuProthGlnAspLeuIleSerVa 212
  |||  |||  |||  |||  |||  |||  |||  |||  |||
621 C...GAAACACATGANTGCGCGCGCGCATCCGCGGCTTTGAGAGCGA 667
  |||  |||  |||  |||  |||  |||  |||  |||  |||
212 lAlaHisLeuHisThrIleAlaGlyProTyrProSerGlySerProSerT 229

```



```

115 saapLeuGluValPheLysLysGluGlyLeuPheAlaLeuPheLysG 132
132 NCGTCGCTTGCAGCAAAATCCCTGCGCGATGCGCGG...CCGTCGCC 444
133 lmrArgProPheasp...lleProAlaLeuProThrGlnSerProHisIle 147
445 ATCTTCGTCATGCGATGACACCAATCCGCTN...GCCGACAGACC 488
148 ValPheIleAsnLeuAlaAspAsnArgProPheThrProSerValIle 164
489 TGTGGTTGTGATCAAGAACCCGNCANAT.....TTCAGAC 526
164 shsLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIlePheVal 181
527 GAGTWTGCTGATGAGCGGTTTGCAGCGAGCTTAACATGCTGTGT 576
181 alclValAlaGlnAlaIleAlaLysLeuPheGlyLeuLysProHisIle 197
577 AAGCAGCTGGCGCAGACGTCGCTGTAATAATGCTGCCACATC...GA 623
198 SerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIleAla 214
624 AACACATGATTCGGCGCGCGCGATCCGCGGTTGAGTGGCAGCACA 673
214 sleuHisThrIleAspGlyProPheProSerGlySerProSerThrHis 231
674 TTCATTTCAATTCAGCCGTC...GGTCAACAAACCGTTGGAGCCATC 720
231 leHisHisIleAlaArgIleAlaGlnArgGlnArgAspAlaIlePheTr 247
721 AATTATCAAGATGTAATTCATCGACGCTGTTGTTGCAACGCGCTGT 770
248 SerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGlyPhe 264
771 GAACACGAGCGCGTATGCTTGGGTGCTTCTCAATC.....AACA 814
264 lleuGlyGlnGluIleValAlaLeuAlaGlySerAlaLeuProProSe 281
815 AACACAGCCTTTCGCTGACCGTTTGGCGAAGTATGCAAAATTA 864
281 lmrArgLysIleLeuIleThrAlaLysGlyAlaSerPheSerPheLeu 297
865 GCGGCGCAATGCGTGGTACGACAGCAAC...CGCGTATTCGCGTTCGT 911
298 ProLysAspIlePheSerSerAspGlnIleThrLeuIleSerGlyAsp 314
912 ATTGAACGCGCGGATTTACACAGCGCGACAT...TATTGGACGCT 958
314 oleuThrGlyArgLeuLysLysGluGlnAsnProCysLeuGlyMet 331
959 ACCACATCAGATTCCTGTTATCGAAGAGCGCGACAGCAAGCGTGT 1008
331 rGAspHisThrIleThrLeuLeuProAsnProLysThrArgIleSerP 347
1009 GCCTGGGTTGCCCGACCGACCAAAATCTCATCAGCGTACGACCT 1058
348 SerPheLeuArgLeuGlyTyrAsnLysLeuThrValThrArgTyrLe 364
1059 CGGCATTTCTCGAAGAAACAA...CTCTCAAGTTTACAGACGCG 1102
364 uSerGlyPhePheLysArgLysArgValPheMetAspMetAspThrAs 381
1103 TCACGCTGGCGACCGCGCATGTCGCGATGTTGACTTACAGCGCGTA 1152
381 eHisGlyGluLysArgProIleIleAspAlaGlnIleTyrGlnArgVal 397
1153 ATGCCGCTAGACATCTGCTACCTGCTTTGCGCGATTTATTCGTCGG 1202
398 SerAlaIleProValProValAlaLeuIleIleLysAlaLeuGluThr 414
1203 CGATACCGACGCGCAAGCATTTGGTGTGATTTGACGAGCAAGAG 1252

```

```

414 nasnPheGluGluAlaCysArgLeuGlyLeuLeuGluValAlaProGlu 431
1253 ACCTGCTTTTGGCAGCTTCCTGCTCCCGCGCAATACGAATANGCCG 1302
431 sPheAlaLeuProThrPheIleAspProSerLysThrGluMetPheSer 447
1303 CTGTCGCTAAGCTGCTGTAACACNTTGAAGAGAA 1338
448 lleValLysGluSerLeuLeuArgTyrAlaLysGlu 459

seq_name: SwissProt_40:RNFC_VIBCH
seq_documentation_block:
ID RNFC_VIBCH STANDARD; PRT; 774 AA.
AC O9KTB8;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfC.
GN RNFC OR VC1015
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=66;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emdolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Niemann W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- CORPORA: Binds 2 4Fe-4S clusters (Potential).
CC -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AEO04183; AAF94176.1; ALR_INT.
DR TIGR: VC1015;
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF01512; Complex1_51K; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane;
KW Complete proteome.
FT METAL 378
FT METAL 381
FT METAL 384
FT METAL 388
FT METAL 417
FT METAL 420
FT METAL 423
FT METAL 427
FT METAL 427
FT SEQUENCE 774 AA; 83500 MM; 44B1E3C304E3A4E CRC64;

```


alignment_scores:

Quality: 157.00 Length: 472
 Ratio: 0.720 Gaps: 18
 Percent similarity: 46.186 Percent identity: 21.822

alignment_block:

US-09-303-518D-127 x RNFC_VIBCH ..

Align seg 1/1 to: RNFC_VIBCH from: 1 to: 774

```

52 CAAGTCATTATGACGGCCGCTCATTAACGAAATCGGCTTGGTGGCA 101
   ||| ||| : : : : : ||| : : : : : ||| : : : : : |||
32 GlnProIleIleAsnAlaSerIleProAsnGluLeuValIleuProLeu 48
   : : : : : ||| : : : : : ||| : : : : : |||
102 ACAAATATCGCGGTATGCGCCCTNGATGAAGTCAAGAGCGCATGCCG 151
   : : : : : ||| : : : : : ||| : : : : : |||
48 sGlnHisIleGlyLysAlaGlyAspLeuLeuValIleGlyAspArgV 65
   : : : : : ||| : : : : : ||| : : : : : |||
152 TCAAAAAGGCCAAGTGTCTTTGAAGACAAAAGATCCGGCGGTGGTG 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 AlenLysIleGlnProLeuThrGlnIleThrSerThrPheMetLeuPro 81
   : : : : : ||| : : : : : ||| : : : : : |||
202 TTTACCGCGCCGTTTACGCAAAATCGCCGCTATCCATCGCGCGAAAA 251
   : : : : : ||| : : : : : ||| : : : : : |||
82 IleHisIleProIleThrSerGlyValIleSerAlaIle.....GluPr 95
   : : : : : ||| : : : : : ||| : : : : : |||
252 GCGCGACTTCAG.....TCGCGTGTGATG 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 oArgThrValAlaHisProSerGlyLeuSerGlyLeuCysIleValLeu 112
   : : : : : ||| : : : : : ||| : : : : : |||
278 CGCTTGAAAGCAACGACGAA.....ATCGAGTTC 306
   : : : : : ||| : : : : : ||| : : : : : |||
112 hrProAsnGlnGlnIleuIleuIleuIleuIleuIleuIleuIleu 128
   : : : : : ||| : : : : : ||| : : : : : |||
307 GAACGCTACGCGCCGACAGCTTGGCAACTTAAGCGCGGANGAANTNG 356
   : : : : : ||| : : : : : ||| : : : : : |||
129 GlnGlnLeuThrProGlnIleuLeuIleuIleuIleuIleuIleuIleu 145
   : : : : : ||| : : : : : ||| : : : : : |||
357 NNGCATTCGATCCATCGGTTTGGACGTGCGTGGCTANCCGTCGT 406
   : : : : : ||| : : : : : ||| : : : : : |||
145 eSerGlyMetGlyGlyAlaGlyPheProThrAla..... 156
   : : : : : ||| : : : : : ||| : : : : : |||
407 TCAGCAAAATCCCTGCGCTCGATGCCGACCGCTGCGCATTCGTGCAT 456
   : : : : : ||| : : : : : ||| : : : : : |||
157 ..LysLysLeuGlnSerGlyLeuSerArgThrGlnIleLeuIleLeu 172
   : : : : : ||| : : : : : ||| : : : : : |||
457 GCGATGACACCAATCCG...CTNCGCGACGACCTGTGTGTGTATCAA 503
   ||| : : : : : ||| : : : : : ||| : : : : : |||
173 AlaAlaGlnCysGlnProIleThrAlaAspValLeuMetArgI 189
   : : : : : ||| : : : : : ||| : : : : : |||
504 AGAAGCCGCGANGATTCAGACGANGTNGCTGTA..... 540
   : : : : : ||| : : : : : ||| : : : : : |||
189 nTyAlaHisGlnIleIleGlnGlyIleGlnIleValGlnHisIleLeu 206
   : : : : : ||| : : : : : ||| : : : : : |||
540 ..... 540
206 ysrProLysLeuThrIleIleGlyIleGluAspAsnLysProGluAlaVal 222
   : : : : : ||| : : : : : ||| : : : : : |||
541 .....TTGAGCCGTTTGACCGACGCTAAATCCATGTGTGAAGCGAC 584
   ||| : : : : : ||| : : : : : ||| : : : : : |||
223 AlaAlaLeuGlnGlnAlaAlaGlnAspLysProMetValIleArgValI 239
   : : : : : ||| : : : : : ||| : : : : : |||
585 TGCGGAGACGTGCGCT.....GAAATGTCTG 613
   : : : : : ||| : : : : : ||| : : : : : |||
239 eProIleLysIleProSerGlyGlyLysGlnIleuIleuIleuIleu 256
   : : : : : ||| : : : : : ||| : : : : : |||
614 CCAACATCGAACAACATGAATTCGCGCGCCGATCCGCGGTTTGACT 663
   : : : : : ||| : : : : : ||| : : : : : |||
256 hrAsnLeuGlnValProLysGlyGlyIleProAlaAspIleGlyLeu... 271
   : : : : : ||| : : : : : ||| : : : : : |||
664 GGCACGCAATTCATTGAGCGGTGCGTGCACAAACAAACCGTTTG 713
   : : : : : ||| : : : : : ||| : : : : : |||

```

```

272 .....MetValGlnAsnValGlySer..... 278
714 GACCATCATATTATCAGATGTAAATTCGACATCGGACGCTTTGTTGCAACG 763
   : : : : : ||| : : : : : ||| : : : : : |||
279 .....LeuGlnAlaIleAlaValGlnAlaIleValHisG 289
   : : : : : ||| : : : : : ||| : : : : : |||
764 GCCGTCTGAACACCGACGCGGTATGCTTTGGGTGCTTCAGTCAAC 813
   : : : : : ||| : : : : : ||| : : : : : |||
289 LysIleProLeuIleArgArgValIleThrLeuThrGlyLysPheGly 305
   : : : : : ||| : : : : : ||| : : : : : |||
814 AAACACAGCCTCTGCGTACACCTTTGGGTGCGCAAGTATCCGAATTAC 863
   : : : : : ||| : : : : : ||| : : : : : |||
306 LysProArgAsnValIleProAlaLeuLeuGlyThrProValGlnAlaLeu 322
   : : : : : ||| : : : : : ||| : : : : : |||
864 TGCGGGCGAATGCTTGACGACGACAAAC.....CCGATGATTCG 904
   : : : : : ||| : : : : : ||| : : : : : |||
322 uAsnGluPheGlyTyrIleLysAlaAspLysLysLeuProArgLeuIleMetG 339
   : : : : : ||| : : : : : ||| : : : : : |||
905 GTTCGGATTGAACGCGCGATTAACACAGCGCGACGATTTTGGGA 954
   : : : : : ||| : : : : : ||| : : : : : |||
339 LysIleProMetMetGlyPheThrLeuPro..... 348
   : : : : : ||| : : : : : ||| : : : : : |||
955 CGCTACCAATGAGATTCGCTATTCGAAAGAGCGCGCAAGAGCT 1004
   : : : : : ||| : : : : : ||| : : : : : |||
349 .....HisAlaGlnValProIle.....ThrLysThrAl 358
   : : : : : ||| : : : : : ||| : : : : : |||
1005 GTTCGGCTGGGTTCGCGCGACCGCGACAAATATCATCATCCGCTACGA 1054
   : : : : : ||| : : : : : ||| : : : : : |||
358 aAsnCysIleLeuAlaPro.....ThrArgAsnG 368
   : : : : : ||| : : : : : ||| : : : : : |||
1055 CCCTCGGCCATTTCCTGAAACAAACTTTCAGTTACAGACGCGCTC 1104
   : : : : : ||| : : : : : ||| : : : : : |||
368 ILeu.....ThrSerSerAsp 373
   : : : : : ||| : : : : : ||| : : : : : |||
1105 AACGGTGCGACGCGCCATGTCGTCGATGTAATTCAGACGCGCTAAT 1154
   : : : : : ||| : : : : : ||| : : : : : |||
374 Asn.....GlnMetAlaCysIleArgCysGlyGlnCysAlaGlnLac 388
   : : : : : ||| : : : : : ||| : : : : : |||
1155 GCGCGTGAACATCTGCTACCTGCTTTGGCGGATTAATGTCGCGCG 1204
   : : : : : ||| : : : : : ||| : : : : : |||
388 sProValSerLeuLeuProGlnGlnIleuGlnIleuIleuIleuIleu 405
   : : : : : ||| : : : : : ||| : : : : : |||
1205 ATACGACACGCGCGCAACGATTCGCTTGGTGTGAATGCGCAAGAAAG 1254
   : : : : : ||| : : : : : ||| : : : : : |||
405 ILeuAspLys.....CysGlnIleuAspLeuLysAsp 416
   : : : : : ||| : : : : : ||| : : : : : |||
1255 .....CTCGCTTGTGACAGCTTGTGTCGCGCGGCAATACGATA 1295
   : : : : : ||| : : : : : ||| : : : : : |||
417 CysIleGlnCysGlyAlaCysAlaTyrValCysProSerGluLeu..... 431
   : : : : : ||| : : : : : ||| : : : : : |||
1296 NGCGCCGCTGTTCGT 1311
   : : : : : ||| : : : : : ||| : : : : : |||
432 .....ProLeuValGln 435
   : : : : : ||| : : : : : ||| : : : : : |||

seq_name: SwissProt_40:HKR1_YEAST

seq_documentation_block:
ID HKR1_YEAST STANDARD; PRT; 1802 AA.
AC P41809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hansenula MRK11 killer toxin-resistant protein 1 precursor.
GN HKR1 OR YDR420W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YNN 295;
RX MEDLINE=94156857; PubMed=8113191;
RA Kasahara S., Yamada H., Mio T., Shitatori Y., Miyamoto C.,

```



```

709 AlavalSerSerThyThyThySerPro.....SerAl 720
321 GGGGCGGTAGCTTCGACATCGATTGCTGCTTCACGCGCATCA 272
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
720 aProAlaAlaIleSerSerThyThyThySerProSerAlaProValA 737
271 CGACCGCATGAAATGACGCGCTTTTCGCCGCGATGATGCGCGATTTTG 222
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
737 laValSerSerThyThyThySerSerProSerAlaProAlaAlaIle... 752
221 CCTGAACNGCGCGGTAAACACGACCGCGGATNCTTTTGTCTTCAAA 172
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
753 .....SerSerThyThyThyThySerSerProSerAlaProValAAlaValSe 767
171 CAGACCTTGCGCTTTTTCAGCGCATCGCTTCCTGACTTTCATTCNAGG 122
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 rSerThyThy.....ThySerSerProSerAlaLeuValValLeuS 781
121 GCGCGATACCGCATATTCT.....TCGCCAAGCAAC 90
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 erSerThySerThySerSerProThyAspIleValThySerProSerThy 797
89 GCGACTTGCGTAATGACGCGC 69
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
798 PheAlaAlaIleSerSerGly 804

```

seq_name: SwissProt_40:YM96_YEAST

seq_documentation_block:
ID YM96_YEAST STANDARD; PRT; 1140 AA.

```

AC 004893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRES-FER4 intergenic region.
GN YMR317W OR YMR924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z54141; CAA90835.1; -.
DR SGD: S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

```

alignment_scores:

Quality: 131.00 Length: 466
Ratio: 0.577 Gaps: 16
Percent Similarity: 48.712 Percent Identity: 22.532

alignment_block:

US-09-303-518d-127/rev x YM96_YEAST ..

Align seg 1/1 to: YM96_YEAST from: 1 to: 1140

```

1331 TCAAGGTTTCAGACCTTACGCAACAGCGCGCCTATTCGTATTGGCC 1282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

386 SerSerThySerSerSerValSerSerGluAlaProSerSer..... 399
1281 CGGGCAGACGAAAGCTGCACAAAGCGAGCTTCTTCGTCAATTCCAAAC 1232
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 .....ThySerSerValSerSerGluA 408
1231 AACCAATGCTTCGCGCTGTCGATGCGCAGCATTAATTCGCGCAA 1182
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 lAProSerSerThySerSerSerValSerSerGluIleSerSerThylys 424
1181 AGCAGGTTAGCAGATGTCGTACGCGCATTCAGCGCTCGTAGTACCAAT 1132
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425 SerSerVal.....MetSerSerGluValSerSerAla.....Th 436
1131 CGGACCAATGCGCGCGTCGCCACCGCTGACGCGCTGCGTGAACCTTGA 1082
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 rSerSerLeuValSerSerGluAlaProSerAlaIleSerSerLeuAlaS 453
1081 GT.....TTGTTTTCAGGAAATGCGCAGGCTGTACGCTGATG... 1041
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 erSerArgLeuPheSerSerLysAsnThySerValThySerThyLeuVal 469
1040 .....GAGTATTTCGCGGTCGCGCGCAACCGCAACCGTTCCTTT 997
470 AlAthrGluAlaSerSerValThySerSerLeuAlaGProSerSerGluTh 486
996 GCTGCGCGCTTCTTCGATACGGAATCGATTGTTGTAGTCGCCAAT 947
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
486 rLeuAlaSerAsnSerIleGlu..... 494
946 AATCTGCGCGCTTGTGTAATCGCGCTTCATATACGAACCGGAATC 897
494 ..... 494
896 AGCGGTTGTCGCTGCAACCAATTCGCCGAGTAATTGGCAATCTTT 847
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 SerSerLeuSerThyThyThySerThyValSerThyThyThySerAl 511
846 CGCACCCAAACGCTACGCAAGAGCGGTGTTGTTGACTTGAAGAACAC 797
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
511 aAlaSerSerThyLeuGlySerlys..... 519
796 CCAAGCAATACGCGCTCGGTTCAGACGCGCTGTGTCAAACAAACGT 747
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 .....ValSerSerSerAsnSerArgMetAlaThySerlys... 531
746 CGCATGCAATTCATCTTGATTAATGATGTCACAAACGGTTTGT... 699
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 ...ThySerSerThySerSerAspLeuSerLysSerSerValIlePheG 547
698 .....GCACCGACCGGCTCATGAATGAATGATCT 671
547 yAsnSerSerThyValThyThySerProSerAlaSerIleSerLeuThA 564
670 GCGTGCACCTCAACGCGCGGATGCGGCGCGCAATTCATGCTTTCG 621
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
564 lAserProLeuProSerValTTPSerAspIleThySerSerGluAlaSer 580
620 ATGTTGACAGATTTTCAGACGCGACGTCGCGCAGCTGCTTACACAC 571
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 SerIleSerSerAsnLeuAlaSerSerSerAlaProSerAspAsnSse 597
570 ATGGATT.....TTACGTCGCTCAACGCGCTCATACACGA 533
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 rThrIleAlaSerAlaSerLeuIleValThyThyThyLysAsnSerVal 614
532 NACNTGTCGAAATCNTGNCGGCTTCTTTCATCAACACAGCGGCT 483
614 alSerSerIleValSerSerIleThySerSerGluThyThyAsnGluSer 630
482 ...GCCGACGAGATGTTGTCATTCGATTGACGAAGATGGCGAAGC 436
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
631 AsnLeuAlaThrSerSerThySerSerLeuLeuSerAsnLysAlaThrAla 647

```



```

706 aThrProlyserGluSerAspAsnProLeuProThrSerSerSerAla 723
      ::::: |||
880 CAACCAATTCGCCGCGAGTAATTTGGCATCTTTCGACCCCAACGTA 831
      ::::: |||
723 laThrThrThrProAla.....SerThrAlaLeuProThrThrAla 736
      ::::: |||
830 CGCAAGAGGCGTGTGTTGTTGACTTGAGAACCCCAAGCAATCAGCG 781
      ::::: |||
737 .....ThrAlaThrAla 740
      ::::: |||
780 CTCGGTGTTCAGAGCGCTGTTGCAACCAACGTCGATGCGCAATTCAT 731
      ::::: |||
740 aHisThrPheLysProIlePheGluSerValGluProPheAla..... 754
      ::::: |||
730 CTGTATTAATGATGTCACCAACGCTTTGTTGACGCGACGCGTCAATG 681
      ::::: |||
755 .....AlaMetProLeuSerProProPheSerLeu 764
      ::::: |||
680 AATGCAATGTCGTCGCCATCAACCGCGCGATGCGCGCGCGAATTC 631
      ::::: |||
765 LysGlnThrThrAlaProIleThrThrAlaThrSerAlaPro..... 779
      ::::: |||
630 ATGTGTTTCAGATGTCGACGATTTTCAGACGACGTCGCGCAGCTG 581
      ::::: |||
780 .....LeuLeuThrGlyLeuGlyThrAlaThrSerThrValAla 793
      ::::: |||
580 CTTACACACATGATTTTACGCTCGTCAACGCGCTCATACACACANA 531
      ::::: |||
793 hr.....GlyThrThrAla 797
      ::::: |||
530 CMTGCTGAATGTCGCGGCTTTGATGACACACACAGGAGTTCG 481
      ::::: |||
798 SerAlaSerLysProValPheGlyPheGlyValThrThrAlaAlaSerTh 814
      ::::: |||
480 CGCAGCGGATGTCGTCATCGCATTCAGACGACGTCGCGCGCTCGG 431
      ::::: |||
814 rAlaSerThrIleAlaSerThrSerGlnSerIleLeuPheGlyGlyAla 831
      ::::: |||
430 CATCAGCGGACGAGGATTTGCTGAACGACGCGATACGACGCGAC 381
      ::::: |||
831 roProValThr.....AlaSerSerSer 838
      ::::: |||
380 AAACCGGATGATGATGATGTCGCGGCTTTGTCGCGCTT...AAGTT 334
      ::::: |||
839 AlaProAlaLeuAlaSerIlePheGlnPheGlyLysProLeuAlaProAl 855
      ::::: |||
333 TGCCAAAGCTTGGCGCGGTAGCGTTCGAACTGATTTGTCGTCGCTT 284
      ::::: |||
855 aAlaSerValAlaGlyThrSerPheSerGlnSerIleAlaSerAlaG 872
      ::::: |||
283 CAACGCGCATCAGACGCGCTGAACTAGCGGCTTTCCGCGGAGGANTG 234
      ::::: |||
872 laThrAlaAlaSerAsn..... 877
      ::::: |||
233 GCGGCGATTTGCTGAACGCGCGGTAAACACGACCGCGGATTC 184
      ::::: |||
878 .....SerSerGlyPheSerGlyPheGlyGlyThrLe 889
      ::::: |||
183 TTTGTCCTCAACAGCACTGCGCTTTTTCGCGCATCGCTTCTCTGA 134
      ::::: |||
889 uThrThrSerThrSerAla...ProAlaThrThrSerGlnProThrLeuT 905
      ::::: |||
133 CTTTATCAAGGCGCGCATACGCGATTCCTTCGCAAGACGCGACT 84
      ::::: |||
905 hrPheSerAsnThrValThrProThrPheAsnIleProPheSerAlaSer 921
      ::::: |||
83 TCGGTAAATGACGCGCGCTCATAAATGACTCTCGCGTCTCGCGCAT 34
      ::::: |||
922 AlaLysProAlaLeuProThrThrTyProGlyAlaAsnSerGlnProThrPh 938
      ::::: |||
33 GGGC 30
      ::::: |||

```

```

938 egly 939
seq_name: SwissProt_40:AMYH_YEAST
seq_documentation_block:
ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucohydrolase).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; Pubmed=3106330;
RT Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STRA1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; Pubmed=3141213;
RX Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STRA2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -I- SIMILARITY: TO S. POMBE SPC215.13.
CC -I- SIMILARITY: SOME, TO S. POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z38061; CAAB6176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; M0C1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
FT SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

```

alignment_scores:
 Quality: 123.00 Length: 532
 Ratio: 0.498 Gaps: 23
 Percent Similarity: 46.429 Percent Identity: 20.489

alignment block:

US-09-303-518d-127/rev x AMYH_YEAST ..

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

```

1337 TCCTCTCTCAANGTTTCCAGCACCTTAGCAACACGGGCGCTATTGCTA 1288
|||||
891 SerPheSerThrGlyThrThrValThrProSerSerLysTyr.... 905
1287 TTTGCCCGGACAGACAGCTGCACAAAGCGCTCTTCCTGTCGCAATT 1238
|||||
906 ...ProGlySerGlnThrGluThrSerValSerSerThrThrGluThr 921
1237 CCAACGACCCCAATGCTGCGCGCTGCGTA.....TCGCGG 1200
|||||
921 hTleValProThrLysThrThrThrSerValThrProSerThrThr 937
1199 AGATTAAATGCGCGCAAGACAGGATGATGCTAGCGGATGATAC 1150
|||||
938 ThrThrThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThr 954
1149 GCGCTCGTAAGTACCAATCGCACCATGGCGCGCTGCGCACGTTGACGG 1100
|||||
954 rThSerGlyCysSerProLysThrValThrThrThrValProThrThr 971
1099 CTGTGTGACGTGAGAGTTGTTTTCAGGAATGGCCGACGGTGTCTA 1050
|||||
971 hThrThrSerValThrThr.....SerSerThrThrThrThrThr 983
1049 GCGGTGATGAGATATTGTCGCGCTGCGCGCAACCCGACGACAGCTC 1000
|||||
984 ThrThrThrValCysSerThrGlyThrAsnSerAlaGlyGluThrThr 1000
999 TTTGCTGCGCGCTTCTGATGATGAGAAATCTGATGTTGTTGAGCTGCCA 950
|||||
1000 rGlyCysSerProLysThrThr..... 1008
949 AATATCGTGGCGCTGTGTATTCGCGCGCTCATACGCAACCGGAA 900
|||||
1009 ...ThrThrValProCysSerThrSerProSerGluThrAlaSerGlu 1023
900 ..... 900
1024 SerThrThrThrSerProThrThrProValThrThrValThrValSerThrThr 1040
899 .....ATCAGCGGCTGTGCTGCTGCAACCAATTCGCCGCACTAATT 857
1040 rValValThrThrThrGluThrSerThrThrLysProGlyGlyGluThr 1057
856 GCGATCTCTTGCACCCAAAGCGTACGCAAGCGCTGCTGTTTACT 807
|||||
1057 hThrThrThrThrValThrLysAsnThr....ProThrThrThrThrThrThr 1072
806 TGAGAACACCCAAAGCAATCAGCGCTGCTGCTGTTACAGCGCGCTGTGC 757
|||||
1073 lLeAlaProThrProSerValThr.....ThrValThr 1083
756 AAACAACACGTCGATGCGCATTTACATTTGATTAATGATGTCCAACGG 707
|||||
1083 rAsnPhrThrProThrThrThrThrThr.....ThyV 1094
706 TTTTGTTCACGACGCGGCTCAATGAATGAATGTCGACACATCAAA 657
|||||
1094 aLysSerThrGlyThrAsnSer.....AlaGlyGluThr 1105
656 CCGGCGGATGCGGCGCGCAATTCATGTTGATGTTGACGATTC 607
|||||

```

```

1106 ThrSerGlyCysSerProLysThrThrValThrThrValProCysSerThr 1122
606 TTCAGACGGC.....ACGTCGCGCACGCTGCTTACACATGCA 566
|||||
1122 rGlyThrGlyGluThrThrThrGluAlaThrThrLeuValThrThrAlaV 1139
565 TTTTACGCTCGTCACCAACGGCTCAATTCACAGCANACNWTGCTGAATCN 516
|||||
1139 alThrThrThrThrValThrThrThrGluSerSerThr.....Gly 1151
515 TCGNCGGCTTCTTGCATCACAACACAGGCTGTCGCGCAACGCGATGCT 466
|||||
1152 ThrAsnSerAlaGlyLysThrThrThrThrGlyTyrThrThrLysSerValPr 1168
465 GTCCATGCGCATTCAGAGATGGCGAAGCGCTGCGCATGACGCGCAGGGA 416
|||||
1168 oThrThrThrValThrThrThrLeuAlaProSerAlaProValThrPro... 1183
415 TTTTGTGAACGAGACGATACGACGCGCAGTCCACAAACCGATTTGATC 366
|||||
1184 .....AlaThrAsnAlaVal.....ProThrThrThr 1192
365 AGATTGCNNMNTNTCTGCGCGCTTACGTTGCCACAGCTTGGCGCGC 316
|||||
1193 ThrThrThrThrGlyCysSerAlaAla.....ThrAsnAlaAlaGly.. 1205
315 GTAGCGTTCGACGCTGATTCGTCG..... 291
1206 ...GluThrThrSerValCysSerAlaLysThrThrLeuValSerSerAla 1221
291 ..... 291
1221 eraAlaGlyGluAsnThrAlaProSerAlaThrThrProValThrThrAla 1237
290 TTGCTTCACGCGCATTCGACGACGACGACTGACG..... 258
|||||
1238 lIleProThrThrValThrThrThrThrGluSerSerValGlyThrAsnSerAl 1254
257 .....ACGCGCTTTGCGCG...CGATGGA 236
1254 aGlyGluThrThrThrGlyTyrThrThrThrLysSerThrProThrThrTyrI 1271
235 TGGCGCGCATTTTTCCT.....GAACNCGC 210
|||||
1271 lThrThrThrLeuThrProGlySerAsnGlyAlaLysAsnTyrGluThrVal 1287
209 GCGGTAAACACCGCGCG.....GGATN 187
|||||
1288 AlaThrAlaThrAsnProLysThrThrThrSerGlnLeuAlaThr 1304
186 CTTTGTGCTTCACACGACGCTGCGCTTTTGAACGCGATGCGCTCT 137
|||||
1304 rThrAlaSerThrThrThrThrValAlaProValThrThr....SerProSerL 1320
136 TGACTTTCAATCNAGGGCGCATACCGGCAATTCCTTCCGCAACGACGG 87
|||||
1320 eThrGly.....ProLeuGlnSerAlaSerGlySerAla 1331
86 ACTTCGTATGACGGCGCGCATTAATGACT.....TCGTC 49
|||||
1332 ValAlaThrThrSerValProSerThrThrThrThrThrThrThrThr 1348
48 CCGTCTGCGCGGATGCGGCGGCTTACAGCTTTTGAATGATTTAAATC 3
|||||
1348 aAsnThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1363

```

seq_name: SwissProt_40:RNFC_PSEAE

seq_documentation_block:

ID RNFC_PSEAE STANDARD; PRT; 774 AA.

AC O9HYB8;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)


```

890 ACCGCGTATTTCCGGTTCGATTTGAACGGCGGATTCACAAAGGCGG 939
      ::::::::::::::::::::
325 luargValleuemetglyly..... 331
940 CACGATTTATTTGGGACGCTACCAATCAGATTCCGTTATGGAAGAG 989
331 ..... 331
990 CCGGCAAAAGAGTGTTCGGTGGGTTGGCCGCGGCGGCAAAATACT 1059
      ::::::::::::::::::::
332 .....PrometMetglyPheAlaLeuProHsp..... 340
1040 CCATCAGCGGTACGACCTCGGCCATTTCCTGAAAACAAACTCTTCAG 1089
      ::::::::::::::::::::
341 .....LeuSerValProLeuIleLys 347
1090 TTCAGACAGCGGTCACAGGTGGCGACCGCGCATGTGCGG..... 1131
      ::::::::::::::::::::
348 ThrCysAsnCysLeuLeuAlaGlyAspAlaThrGluLeuProGluProVa 364
1132 .....ATTGTACTTACGAGCGGCTAATGCGCGC 1159
364 lProAlaMetProCysIleAlaGlyCysGlyAspCysAlaGlnValCysProV 381
1160 TAGACATCTTCCTACCTGCTTTGGCGGATTAATCGTCGCGGATACCC 1209
      ::::::::::::::::::::
381 AlSerLeuLeuProGln...GlnLeuHisPhePheAlaLeuGlyAsp... 395
1210 GACAGCGCGCAAGCATTTGGTGGTTCGTAATGGACAGAAAGACCTGGC 1259
      ::::::::::::::::::::
396 GluHisGluGlnLeuLeuAlaHisAsnLeuPheAspCysIleGlyCysG1 412
1260 TTTGTGACGTTCTGCTGCGCGGCAATACGATTAATGCGCGCTGTTCG 1309
      ::::::::::::::::::::
412 yAlaCysAlaThrValCysProSerSerIle.....ProLeuValG 426
1310 GT 1311
      ::
426 In 426
seq_name: SwissProt_40:VGLX_HSVB
seq_documentation_block:
ID VGLX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN 111
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M86664; AAB02506.1;
CC PIR: H36802; VGBEX1.
CC Glycoprotein; Transmembrane; Signal.
CC SIGNAL 1 22 POTENTIAL.

```

```

FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;

alignment_scores:
  Quality: 116.50 Length: 450
  Ratio: 0.566 Gaps: 15
  Percent Similarity: 45.778 Percent Identity: 21.111

alignment_block:
US-09-303-518d-127/rev x VGLX_HSVB ..

Align seg 1/1 to: VGLX_HSVB from: 1 to: 797

1322 TCCAGACCTTACGCAACAGCGGGCCNTATTCGATTTGCCCCGCGAC 1273
      ::::::::::::::::::::
50 SerSerSerProThrThrSerProThrThrSerSerSerProThr 66
1272 GAAGCTGCACAAAGCAGAGTCTTCTTCGCAATTCACAAACCAATG 1223
      ::::::::::::::::::::
66 rSerThrHisThrSerSerProSerSerThrThrGlnSerSerSer 83
1222 CTTCGCGCGCTGTGTATG...CCGACGATTAATCGGCCAAAGCAGG 1176
      ::::::::::::::::::::
83 hValAlaThrSerSerSerAlaProSerThrAlaSerThrThrSer 99
1175 GTAGCGCAGATGCTACGCGCATTAACGCTGTAGTACCAATGCGCAC 1126
      ::::::::::::::::::::
100 IleProThrSerThrSerThrGluThrThrThrThrPro....Th 114
1125 CATGCGCGGTCGCGCACGCTGACGGCTGCTGAACCTTGAAGATTGT 1076
      ::::::::::::::::::::
114 rAlaSerThrThrProThrThrThrAlaAla..... 126
1075 TTTTCAGAAATGCGCGAGGTCGTACGCGTGATGAGATTGTCGCGC 1026
      ::::::::::::::::::::
127 .....ProThrThrAlaAlaThrThrThrAlaValAlaThrThr 138
1025 TCGCGCGCAACCGCGGCAACAGCTTTGCTGCGGCTCTTCGATAC 976
      ::::::::::::::::::::
139 AlaAlaSerThrSerAlaGluThrThr..... 148
975 GGAATCTGATTTGTGTAGCTGCCAATAATATCGCGCGCTGTGTAA 926
      ::::::::::::::::::::
149 .....AlaThrAlaThrAlaThrS 155
925 TCGCGCGCTTCAATNCCGAACCGGAATCAGCGGTGTCTGCTCAAC 876
      ::::::::::::::::::::
155 eThrProThrThrThrThrProThrSerThrThrThrThrAlaThr 171
875 AATTCGCCCGCAGTATTTGCGATCTTCGACCCAAACGCTGCGCA 826
      ::::::::::::::::::::
172 .....ThrThrValProThrThrAlaSerThr 180
825 GAGCGCTGTTGTGTGACTTGAAACCAACCAACCAATCAGCGCTGG 776
      ::::::::::::::::::::
180 rThrThrAspThrThrThrAlaAlaThrThrThrAlaAlaThrThrTha 197
775 TGGTACAGAGCGCTGTGCAACAAACGTCGATGGCAATTAACATCTGA 726
      ::::::::::::::::::::
197 lAlaAlaThrThrAlaAlaThrThrThrAlaAlaThrThrThrAla 213
725 TAATGATGTCCAACAGGTTTGTTCACAGCAGCGCTCAATGAATG 676
      ::::::::::::::::::::
214 ThrThrThrAlaAlaThrThrThrAlaAlaThrThrSerSer..... 227
675 AATGTGCGTGCCATCAACAGCGCGGATGCGGCGCGCGCAATTCATGTG 626
227 .....

```


FT CARBOHYD 533 533 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 886 886 N-LINKED (GLCNAC...) (POTENTIAL)
 SQ SEQUENCE 1117 AA; 121063 MW; 87F13A07E42B0AD1 CRC64;

alignment_scores:

Quality: 110.00 Length: 309
 Ratio: 0.688 Gaps: 12
 Percent Similarity: 51.780 Percent Identity: 22.977

alignment_block:

us-09-303-518d-127/rev x YN96_YEAST

Align seg 1/1 to: YN96_YEAST from: 1 to: 1117

```

960 GTAGCGTCCCAAAATATCGTGGCGCTTGTGTATCGCGCGCTTCAATA 911
|||||.....|
50 ValAlaGlnSerValValVal.....ThrArgVal 59
910 CCGAACCGGAATACAGCGGCTTGTCTGCTC.....AACCAATCG 879
|||||.....|
59 lserAlaSerHisSerProLeuSerValSerProLysAsnArgValS 76
869 CCGGAGTAATTTGGCATACTTTCGACCCAAACGATGCGAGAGGCG 820
|||.....|
76 erAlaSerProLeuSerGlnAspSerAspSerValThrArgThrala 92
819 TGGTTGTGACTTGAGAACACCCAAAGCAATACGCGCTCGGTGTCA 770
|||||.....|
93 .....ValGlnLeuSerLeuSerLeuSerValAlaSerGlnValS 107
769 GACGGCGCTTGCACAAACGCTCGGATGCGCAATTCATCTGTATATG 720
|||.....|
107 roLInylSerAlaGlnThrAsnAsn..... 116
719 ATGCTCCAAACGCTTGTTCACACCGCGCTCATGAAATGATG 670
|||.....|
117 .....AspProValThrValSerAsnLe 124
669 CGTGCACACTCAACCGCGCGGATGCGGCGCGCAATTCATGTTTGA 620
|||||.....|
125 TTTAlaAsnAspAsnSerSerSerSerValHisAsnLeuSerSe 141
619 TGTGGCACCATTTCAGACGCGCAGCTGCGCAGCGCTTACACACA 570
|||||.....|
141 rValSerGlyValAlaSerValMetProSerAlaSerThrMetArgLysV 158
569 TGGATTTCAGCTCGGTCAACGCTCAATACACAGCANACNTGCTGANA 520
|||.....|
158 alThrThrLeuLeuSerGlnThrAla.SerThrSerThrSerThrLeuPh 174
519 ATCNTGCGCGCTTCTTGATC.....ACAACCACAG 488
|||||.....|
174 eSerSerSerLeuSerLeSerGlyThrGlnLeuAsnGlnGlyThrLeuLeuT 191
487 GGTCTCCCGCCNAGCGATGTTGTCATGCAATGACAGAGATGCGGCAAC 438
|||||.....|
191 hSerValSerSerGlyThrLeaSerProLeuValThrGlnMetProSer 207
437 GGTCTCGCATCGACGCGAGGATTTTGTGAACGAGGATNACGACGCC 388
|||||.....|
208 TyrSerSerGlnGlnThrLysIleIleProSerSerLeuThrSerAsnLys 224
|||||.....|
224 sTrh.....leryThrIleSerValArgThra 234
337 AGTTTCCCAACGCTTGGCGCGGTAGCTTCGAACATCGATTCATGCTGTG 288
|||.....|
234 snAlaIleThrAlaThrArgLys.....GluAspSerPheIleAlaSerThr 248
287 CCTTCAACGCAATTCACGACCGACTGAATACCGCGCTTTTCCCGCGATG 238
|||||.....|

```

```

249 ProAlaSer.....SerThrLeuPheTyrProSerAs 259
237 GATGGCGCGCATTTTCCGTAACACNGCGCGGTAAACACGACCGCGAT 188
|||.....|
259 nSerThrGlnAspLeuValGlnThrLeuAla...SerThrThrAlaSerP 275
187 NCTTTTGTCTTCAACACGACTTGGCTTTTGTGACGCGATCGCTTGC 138
|||||.....|
275 roAlaTyrProSerAsnArgThrGln.....lIerThrLeuSerProSer 289
137 TTGACTTTCATCNAGGCGCGCATACCGGCATATCTTCCGCAACGACGC 88
|||||.....|
290 ValSerLeuTyrSerThrThrSerProIleTyr.....ProSerAsnIl 304
87 GACTCGGTATACGCGCGCGCTCA 63
|||||.....|
304 eThrGlnAsnGlySerSerProSer 312
seq_name: SwissProt_40:MUFE_HAEIN
seq_documentation_block:
ID MUFE_HAEIN STANDARD: PRF; 457 AA.
AC P45061:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoyl-alanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MUFE OR H1134.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminopimelate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE MURCDE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32793; AAC22789.1; -.
CC HSRP: P11880; IG64.
CC
CC TIGR: H11134; -.

```

670 CGAATG.....GTTCACGCAGACAACC 892

[illegible]


```

893 GCGTATTTCCGGTTCGGTATTTGAACGGCGCATTTACACAGCGCCGAC 942
      |||::: |||::: |||
339 rgleuleuleuglyglypromelethymetproleleln..... 352
      |||::: |||::: |||
943 GATTATTTGGAGCGCTACCAATCAGATTTCCTTATTCGAGAGAGCGG 992
      |||::: |||::: |||
353 .....AsnProArgValProValIalysgly..... 361
      |||::: |||::: |||
993 CAGCAAGAGCTGTTCCGCTGGGTTGGCGCCGACCGACAAATACATCA 1042
      |||::: |||::: |||
361 ..... 361
1043 TCACGCGTAGACCCCTGCGCATTTCTGAAAAAACAATCTCAAGTTC 1092
      |||::: |||::: |||
362 .....ThraAnglyleleulaleu 368
      |||::: |||::: |||
1093 ACGACAGCCGTCACAGCGTGGCGCCGCGCATG.....GTGCGGATTGG 1136
      |||::: |||::: |||
369 ThrIalalaglutnProglualalysThmetProcstleargysgl 385
      |||::: |||::: |||
1137 TACTTAGAGCGCGTAATGCCGTAGACATCTGCTACCTGCTTTTGC 1186
      |||::: |||::: |||
385 YArgCysValGlnGlyCysProValGlyLeuThProhegluleuasna 402
      |||::: |||::: |||
1187 GCGATTATATGTCGCGCATACGACAGCGCGCAAGCATTTGGCT..... 1230
      |||::: |||::: |||
402 laarglyleHisIalaglyaspLeuGlnGlyIalalysValGlyLeuMet 418
      |||::: |||::: |||
1231 ..TGCTTGAATTTGGACGAAGAAGACCTGCTTGTGACGCTTCGCTG 1277
      |||::: |||::: |||
419 AspycIleuala.....CysGlyCysCysSeryIasncy 430
      |||::: |||::: |||
1278 CCGGCGCAATATGCAATANGCGCCGCTGTTGCGT 1311
      |||::: |||::: |||
430 sprolaasneleu.....ProleuValGln 438
      |||::: |||::: |||
seq_name: SwissProt_40:YGY3_HALSQ
seq_documentation_block:
ID YGY3_HALSQ STANDARD; PRT; 437 AA.
AC P21561;
RX 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
DE 3).
OS Haloferax sp. (strain Aa 2.2).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91100352; PubMed=1846146;
RA Holmes M.L., Dyall-Smith M.L.;
RT "Mutations in DNA gyrase result in novobiocin resistance in
RT halophilic archaeobacteria."
RL J. Bacteriol. 173:642-648(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38373; -; NOT_ANNOTATED_CDS.
DR PIR: C39135; C39135.
KW Hypothetical protein.
SQ SEQUENCE 437 AA; 50626 MW; B5B99A2AF3892BEF CRC64;

```

alignment_scores: Quality: 108.00 Length: 480

Ratio: 0.554 Gaps: 27
Percent Similarity: 40.625 Percent Identity: 23.958

alignment_block:

us-09-303-518d-127 x YGY3_HALSQ ..

Align seg 1/1 to: YGY3_HALSQ from: 1 to: 437

```

12 CAAAAAAGGCTTAACCTGCCATCGCGGCGACGCGGACAAATGATTT 61
      |||::: |||::: |||
48 GlnaspAlaArglyspProAlaArgValProGlnleu..... 62
      |||::: |||::: |||
62 ATGACGCGCGCTCATACGAGATGCGGTGGTGGCGAAGATATGCG 111
      |||::: |||::: |||
63 ...ArgGlyArgAspPheAlaLeuArgArgAlaAspArgVal... 77
      |||::: |||::: |||
112 GGTATGCGCGCCCTNGATGAAGTGAAGAGAGCGATGCCGTAAGAAAG 161
      |||::: |||::: |||
77 LnhIstValProleu.....ArgGlyArgHisProArgValArgArg 90
      |||::: |||::: |||
162 ...CCAAGTCT.....GTTTGAAG 178
      |||::: |||::: |||
91 ValProGlnArgAspGlnAspGlyAlaProArgArgHisLeuLeuAr 107
      |||::: |||::: |||
179 ACAAAAGNATCCGGG...CGTGCTTTAACCGC..... 209
      |||::: |||::: |||
107 garArgValGlyGlyHisArgGlyArgAsnArgHisAlaGlyAspArg 124
      |||::: |||::: |||
210 .....GCCNGTTTCAGGCAAAATCGCGCGCATTCATCG 242
      |||::: |||::: |||
124 rgaIaProGlyAlaAspSerArgLeuArgGlnGlnHisGlnHisProArg 140
      |||::: |||::: |||
243 CCGGCGAAAGCGGCTACTGACGCGTGCATGTCGCGCTTGAAGCAAGC 292
      |||::: |||::: |||
141 GlyArgHisAlaSerAspArgValGlnAspGlyAla.....HisProAr 155
      |||::: |||::: |||
293 ACGAAATCGAGTTGCAACGCTACGCGCCGCAAGCGTTGGCAAACTTAAC 342
      |||::: |||::: |||
155 garGlnArgLeuArgGln...GlnProArgHisAlaGlyArgProArg 171
      |||::: |||::: |||
343 GCGGANGAANTNNGNCAATGATCAATCGGTTGTGATGCGCT 392
      |||::: |||::: |||
171 rgarArg..... 172
      |||::: |||::: |||
393 GCGTACCGCTCGTTAGCAAAATCCCTCGCTGCATGCCGAGCC..... 437
      |||::: |||::: |||
173 .....GlnProProArgArgGlyArgSerArgGln 182
      |||::: |||::: |||
438 .....GTTCCGATCTTCGTCATCGATGCGACCAATCCGCTNGCG 480
      |||::: |||::: |||
182 yThrHisArgArgHisLeuArgGlnAlaPro..... 192
      |||::: |||::: |||
481 GCAGACCGCTGT.....GGTTGATCAAAAGAGCGCGANGATTTACG 524
      |||::: |||::: |||
193 ..ArgProAlaValArgGlyProAspGlyAspGlnAlaArgGlnupheArg 208
      |||::: |||::: |||
525 ACGANGTNTGCTGTATTTGAGCCGTTTGACCGAGCGTAAATCCATGTGT 574
      |||::: |||::: |||
208 ..... 208
      |||::: |||::: |||
575 GTAAGCAGCTGGCGCAGACGTCGCTGTAATAATGCTGCCAATCGAA 624
      |||::: |||::: |||
209 ...GlyProArgHisArgArgGln..... 215
      |||::: |||::: |||
625 ACACATGAATTCGGCGCGCGCATCGCGCGG.....TTT 659
      |||::: |||::: |||
216 .....ArgHisProProThrAlaArgAspValLeuArgGlyGln 228
      |||::: |||::: |||
660 GAGTGCACAGCACATTCATTCATGAGCCGTCGCTGCAAAACAACCG 709
      |||::: |||::: |||
229 ProGlyHisGlyAspGlyHisLeuGlnGlyArgArgGlyArgProAr 245
      |||::: |||::: |||
710 TTTGACCATCAATTAATCAAGATGTAATTCATGCGAGCGTTGTTGCA 759

```



```

491 ACAGGCTGCGCCGACGATGGTGTCCATCGCATTTGACGAAAGATGC 442
      ::::::::::: ||| :::::::::::
68 SerSerSerLeuThrSerThrSerThrSerLeuValSerHisTh 84
      ::::::::::: ||| :::::::::::
441 GACAGGCTGCGCCGACGATGGTGTCCATCGCATTTGACGAAAGATGC 392
      ::::::::::: ||| :::::::::::
84 rSerSerSerThrSerLeuValSerThrSerThrSerThrSerThr 97
      ::::::::::: ||| :::::::::::
391 GCGCAGTCCACAAACCGGATTTGATGATGATGATGATGATGATG 342
      ::::::::::: ||| :::::::::::
98 ..... PhSerPhSerPhSerPhSerPhSerPhSerPhSerPh 106
      ::::::::::: ||| :::::::::::
341 CTTAAGTTGGCCACGCTTGGCCGCGCTTGGATGATGATGATGATG 292
      ::::::::::: ||| :::::::::::
107 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 123
      ::::::::::: ||| :::::::::::
291 GTTGCTTCACAGCGATCAGACGCGATGATGATGATGATGATG 242
      ::::::::::: ||| :::::::::::
123 rLeSerSerSerSerSerSerSerSerSerSerSerSerSerSer 140
      ::::::::::: ||| :::::::::::
241 GATGATGCGCGCGCTTGGCTGAACGCGCGGTAAACACGACGCG 192
      ::::::::::: ||| :::::::::::
140 hrSerThrSerSer.....SerSerSerLeuSerSerThrPro 152
      ::::::::::: ||| :::::::::::
191 GGATNCTTTTGTCTTCAACAGCAGCTTGGCTTTTGGACGCGATG 142
      ::::::::::: ||| :::::::::::
153 .....SerSerSerSerSerSerSerSerSerThrThrThr 165
      ::::::::::: ||| :::::::::::
141 TTCCTGACTTTCATCNAAGGCGCATACCGGATTCCTTGGCAACA 92
      ::::::::::: ||| :::::::::::
165 oSer.....ThrSerSerThrProSerThrThrAlaThrAla 181
      ::::::::::: ||| :::::::::::
91 ACAGGCTGCGCCGATGATGATGATGATGATGATGATGATGATG 69
      ::::::::::: ||| :::::::::::
181 hrLeThrSerLeuThrLeuHisnly 188
      ::::::::::: ||| :::::::::::
seq_name: SwissProt_40:AGAL_YEAST

seq_documentation_block:
ID AGAL_YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-agglutinin attachment subunit precursor.
GN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RL MEDLINE=91304412; PubMed=2072914;
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT "The AGAL product is involved in cell surface attachment of the
RT Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RL Mol. Cell. Biol. 11:4196-4206(1991).
RN [2]
RP SEQUENCE FROM N.A.
RL Pohl T.M.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S-CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M60590; AAA34382.1; -
DR EMBL: 271659; CAN96325.1; -
DR PIR: S17031; S17031.
DR PIR: A41258; A41258.
DR SGD: S0005327; AGAL.
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
KW Phormone response.
FT SIGNAL 1 22
FT CHAIN 23 725
FT DOMAIN 53 493
FT REPEAT 53 149
FT REPEAT 395 493
FT REPEAT 182 307
FT FT 182 307
FT REPEAT 182 188
FT REPEAT 189 195
FT REPEAT 196 202
FT REPEAT 203 209
FT REPEAT 210 216
FT REPEAT 217 223
FT REPEAT 224 230
FT REPEAT 237 244
FT REPEAT 238 244
FT REPEAT 245 251
FT REPEAT 252 258
FT REPEAT 259 265
FT REPEAT 266 272
FT REPEAT 273 279
FT REPEAT 280 286
FT REPEAT 287 293
FT REPEAT 294 300
FT REPEAT 301 307
FT REPEAT 307 307
SQ SEQUENCE 725 AA; 73353 MW; 70420C853B0B01P8 CRC64;

alignment_scores:
Quality: 107.50 Length: 254
Ratio: 0.846 Gaps: 11
Percent Similarity: 50.000 Percent Identity: 24.803

alignment_block:
US-09-303-518D-127/rev x AGAL_YEAST ..
Align seg 1/1 to: AGAL_YEAST from: 1 to: 725
618 GTTGGCAGCATTTTTCAGACGACGCTGCGCGATGCTTACACAT 569
      ||| ||| ||| ::::::::::: ||| |||
124 ValThrSerLeuThrSerThrSerThrSerLeuValSerHisTh 139
      ::::::::::: ||| :::::::::::
568 GGAATTTTACGCTGCGTCAACAGCGCTCAATAC.....AGANA 531
      ::::::::::: ||| :::::::::::
139 hrAlaLeuSerSerLeuSerGluValGlyThrThrValValSerSer 155
      ::::::::::: ||| :::::::::::
530 CNTGCTGAATATCMTGCGCGGCTTGTGATC..... 498
      ::::::::::: ||| :::::::::::
156 SerAlaIleGluProSerSerAlaSerIleIleSerProValThrSerTh 172
      ::::::::::: ||| :::::::::::
497 .....ACAA 494
      ::::::::::: ||| :::::::::::
172 rLeuSerSerThrThrSerSerSerAnProThrThrThrSerLeuSerSer 189
      ::::::::::: ||| :::::::::::
493 CCACAGGCTGCGCCGACGATGGTGTCCATCGCATTTGACGAAAGATG 444
      ::::::::::: ||| :::::::::::
189 hrSerThrSerProSerSerThrThrThrSerProSerSerThrSerThr 205
      ::::::::::: ||| :::::::::::

```

```

443 GCGAAGCGCTCGCATCGACGCGACGAGATTTGCTGACGACGACGAGTACG 394
      :::::::::::::::::::::
206 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSe 222
      :::::::::::::::::::::
393 CAGCGCATCCACAAACCGATGGATCAGA.....TTGC 359
      :::::::::::::::::::::
222 rSerSerSerSerSerSerSerSerSerSerSerSerSerSer 239
      :::::::::::::::::::::
358 NNCNNANNTCTMTCGCGCTTAAGTTGGCAACGCTCGGGCGGAGGCT 309
      :::::::::::::::::::::
239 eTThSerSerSerSerSerSerSerSerSerSerSerSerSer 255
      :::::::::::::::::::::
308 TCGAAGCTGATTCGCTGCTGCTTCAACGCGCAATCAGACGACGAG 259
      :::::::::::::::::::::
256 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSe 272
      :::::::::::::::::::::
258 TACGCGCTTTTCGCGCGCATGAGGCGGCGATTTGCTGAAACNGGCG 209
      :::::::::::::::::::::
272 rThSerSerSerSer.....SerSerSerSerSera 284
      :::::::::::::::::::::
208 CGGTAACACACGCGCGGATTCCTTGTCTTCAACGACGCTGGGCT 159
      :::::::::::::::::::::
284 lASerSerSerSerSerSerSerSerSerSerSerSerSerPro 298
      :::::::::::::::::::::
158 TTTTTCAGCGCA.....TCGCGCTGCTTGAC 133
      :::::::::::::::::::::
299 SerLeuThSerSerSerSerProThLeuAlaSerThSerProSer 315
      :::::::::::::::::::::
132 TTTTCATCNAGGCGGCAT..... 114
      :::::::::::::::::::::
315 rSerLeSerSerThrPheThrAspSerThSerSerLeuAlaSer 332
      :::::::::::::::::::::
113 .....CCGCA 108
      :::::::::::::::::::::
332 lAlaSerSerSerThrSerValSerLeuThSerSerSerThProVal 348
      :::::::::::::::::::::
107 TATCTTCGCGCAACGACGCGATTCGCTGATGACGCGCGCTCAAT 58
      :::::::::::::::::::::
349 rYSerValProSerThrSerSerValAlaThr...ProSer...Me 363
      :::::::::::::::::::::
57 GACTTGCTCC 48
      :::::::::::::::::::::
363 tThSerSer 366
      :::::::::::::::::::::
seq_name: SwissProt_40:N121_HUMAN
seq_documentation_block:
ID N121_HUMAN STANDARD: PRT: 1229 AA.
AC Q9Y2N3; 075115; Q9Y4S7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
DE of 121 kDa) (P145).
GN NUP121 OR KIA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP Cordeas M., Bauer C., Holmes A.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE=Brain;
RC MEDLINE=9840380; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).

```

```

RN [3]
RP SEQUENCE OF 1130-1229 FROM N.A.
RC TISSUE=Uterus;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
CC REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
CC OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
CC MEMBRANE (BY SIMILARITY).
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC006014; AAD28064.1; -
DR EMBL; AB014518; BAA31593.1; -
DR EMBL; AL080109; CAB45713.1; -
KW Nuclear protein; Transport; Transmembrane; Repeat.
FT DOMAIN 1 40
FT TRANSMEM 41 61
FT DOMAIN 62 1229
FT DOMAIN 4 10
FT DOMAIN 51 56
FT DOMAIN 294 299
FT DOMAIN 441 444
FT DOMAIN 499 502
FT DOMAIN 733 736
FT DOMAIN 819 826
FT DOMAIN 869 875
FT DOMAIN 1061 1067
FT DOMAIN 297 297
FT CONFLICT 336 336
FT CONFLICT 379 379
FT CONFLICT 451 451
FT CONFLICT 545 545
FT CONFLICT 559 559
FT CONFLICT 702 702
FT CONFLICT 881 881
FT CONFLICT 965 965
FT CONFLICT 982 984
FT CONFLICT 991 996
FT CONFLICT 1000 1000
FT CONFLICT 1006 1008
FT CONFLICT 1044 1044
FT CONFLICT 1102 1102
FT CONFLICT 1124 1124
FT CONFLICT 1146 1146
FT CONFLICT 1165 1165
FT CONFLICT 1195 1195
FT CONFLICT 1199 1229
SQ SEQUENCE 1202 1202 AA; 125087 MW; 515655D1285898B CRC64;

```

alignment_scores: Quality: 107.00 Length: 425
Ratio: 0.546 Gaps: 17
Percent Similarity: 46.118 Percent Identity: 22.824

alignment_block: US-09-303-518D-127/rev x N121_HUMAN ..

Align seg 1/1 to: M121_HUMAN from: 1 to: 1229

```

1115 TCGCCACCGCTTGAGCGCTGCTGTAAC.....TTGAGAG 1081
      |||||
561 SerProthSerThraLeuAlaProSerThraSnpProLeuLeuGlu 577
1080 TTTGTTTTCAGGAATGCGCAGAGGTCGTACGCGTGAGTATTTGT 1031
      |||||
577 rLeuLysLysMetGlnThrProSerSerLeuProCysProGluSerA 594
1030 CCGGCTCGCGGCAACCCAGCCAGCAGC.....TCGTTG 996
      |||||
594 IaGlyAlaAlaThrThrGlnAlaLeuSerProProLysThrProSerLeu 610
995 CTGGCGCCTTCTTGATATACGAATCTGATGTGTAGCGTCCCAATA 946
      |||||
611 LeuProProLeuGlyLeuSerGln..... 618
945 ATCGTGGCGCCTTGTGTATTCGCGCG.....TTCAATACGGAAC 905
      |||||
619 ..SerGlyProProGlyLeuLeuProSerProSerPheAspSerLysP 634
904 CGGAATACAGCGGTTGCT.....CGCTCAACCAATTGCGCCGCA 864
      |||||
634 ProProThrThrLeuLeuGlyLeuLeuProAlaProSerMetValProAla 650
863 GTATATTTCGATCTTTCGACCCCAAAACGTCAGCAAGAGCGGTGTTT 814
      |||||
651 Thr.....AspThrLysAlaProProThrLeuGlnAlaGlnThrAla 665
813 GTTGACTTGGAACA.....CCCAAGCAATACGCGCT 779
      |||||
665 r.....LysProGlnAlaThrSerAlaProSerProAlaProLysC 679
778 CGGTGTTTCAGACGCGCTTGCAACAACAGTCCGATG..... 741
      |||||
679 InSerPheLeuPheGlyThrGlnAsnThrSerProSerSerProAlaAla 695
740 ...GCAATTACATCTTGATTAATGATGTCACAAACGTTTGTGTCACC 694
      |||||
696 ProAlaAlaSerSerAlaSerProMetPheLysProLeuThrAlaPro 712
693 GACCGGCTCATGAATGAATGTCGTCGCACTCAAAACG..... 654
      |||||
712 opProLysSerGlnLysGlnGlyProThrProGlyProSerValThrA 729
654 ..... 654
729 IaThrAlaProSerSerSerLeuProThrThrSerThrAla 745
653 .....GCCGATGCGGCGCGCGCAATTCATGATGT 625
      |||||
746 ProThrPheGlnProValPheSerSerMetGlyProProAlaSer...Va 761
624 TTGATGTTGGCAGCATTTTCACAGCGCAGCTGGCCACGCTGCTTAC 575
      |||||
761 lProLeuProAlaProPhePheLysGlnThrThrProAlaThrAla 778
574 ACACATGATTTTACGTCGCTCAACAGCGCTCAATACGACANACTGCT 525
      |||||
778 rothr..... 779
524 CTGAATACNTGNCGGCTCTTTGATACACACAGGCTGCGCCGNA 475
      |||||
780 .....ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSe 793
474 CGGATGTTGTCATGCGATGACGAAGATGGGAACGCTGCGCATC 426
      |||||
793 rAlaValAlaProLeuThrSerAlaSerProSerThrAspSerAlaSerL 810
425 ....ACGGCAGGATTTGCTGACGACGAGGTACGACGCGACATCAC 381
      |||||
810 ysrProAlaPheGlyPheGlyLysAsnSerValSerSerSerVal... 825

```

```

380 AAACCGGATTGATGATGATGCGNNNNNTGTCGCCGCTTAAGTTGC 331
      ::
826 SerThrThrThrSerThrAlaThrAlaAlaSerGlnProPheLeuPheG 842
330 C.....AACGCTCGGGCGCGCTGACGCTTCGAACTCGATTCTGCTG 288
      ::
842 yAlaProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerL 859
287 .....CCTTCACGCAATCAGCAGCACTGAAGTACGCC 252
859 heGlnPheGlyLysProProAlaLeuProThrThrThrValThrThr 875
251 TTTTCGCCGCGATGATGCGCGCATTTTGCCTGAACNGCGCGGTAA 202
876 PheSerGlnSerLeuProThrAlaValProThrAlaThrSerSerAl 892
201 CACACGCGCGGATGCTTTGCTTCAACAGCACT...TGCGCTTTT 155
892 aAlaAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAl 909
154 TGACGCGATCGCTTCTTGACTTTCATCAGGCGCATACCGGATAT 105
909 hrSerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrP 925
104 TCTTCGCCACAGCAGCGCATCTCGTAAATGACGCGCGCTCAATATG 55
926 AsnThrPheGlySerSerAlaLysSerProLeuProSerTyProG 942
54 TTGCTCGGCTCGCCCGGATGGC 30
942 yAlaAsnProGlnProAlaPheGly 950

```

seq_name: SwissProt.40:ATPB_PYLII

seq_documentation_block:

ID ATPB_PYLII STANDARD; PRT; 481 AA.

AC P26532;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DE 01-MAR-2002 (Rel. 41, Last annotation update)

DE ATP synthase beta chain (EC 3.6.3.14).

GN ATPB.

OS Pyralia litoralis.

OG Chloroplast.

OC Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;

OC Acinetosporaceae; Pyralia.

OX NCBI_TaxID=2885;

RN [1]

SEQUENCE FROM N.A.

RA MEDLINE=92216062; PubMed=1532750;

RT Jouanin S., Kerbourc'H C., Kloareg B., Loiseaux-De Goer S.;

RT "Nucleotide sequences of the atpb and the atpe genes of the brown

algae Pyralia litoralis (L.) Kjellm.";

RL Plant Mol. Biol. 18:819-822(1992).

CC -I- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON

CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC

CC SUBUNIT

CC -I- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +

CC H(+)(Out).

CC -I- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC

CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE

CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)

CC HAS THREE MAIN SUBUNITS: A, B AND C.

CC -I- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.

CC -I- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

FT CARBOHYD 104 104 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 106 106 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 107 107 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 108 108 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 110 110 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 114 114 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 117 117 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 123 123 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 124 124 O-LINKED (GALNAC. . .) (MUCIN TYPE)
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 917 917 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 985 985 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL)
SO SEQUENCE 1150 AA; 109615 MW; 3CB6B5D29DD7F5A CMC64;

alignment_scores:
Quality: 105.00 Length: 450
Ratio: 0.507 Gaps: 22
Percent Similarity: 46.000 Percent Identity: 24.667

alignment_block:

US-09-303-518D-127/rev x APMU_PIG

Align seg 1/1 to: APMU_PIG from: 1 to: 1150

1325 GTTCCACACCTTACGACAGCGGCGCCTTATTCGTTGCGCGCA 1276
|||||:||||| ||| ||| |||
242 VAlThrGluThAlaArgProSerValAlaGlySerGlyThrGly.. 257
1275 GACGAGCTGACACAGAGAGTCTTCGTCACATTCGACCAAGCA 1226
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
258 .ThrValSerGlyAlaSerGlySerThrylSerSerSerGlySerPro 274
1225 ATGCTGGCGCGCTGCGTATCGCCGACGATTAATCGCGCAAGACAG 1176
|||||: |||: |||: |||: |||: |||: |||: |||: |||
274 lAlaThrGlyAlaSerIleGlyInProGluThrSerArgIleSerVal 290
1175 GTAGCG.....AGATGTCAGCGGCAATTCAGCGCTGCTGA 1141
|||||: |||: |||: |||: |||: |||: |||: |||: |||
291 AlaglySerSerGlyAlaProAlaValSerSerGlyAlaSerGlnAla.. 306
1140 AGTACCAATCGCACCATGCGCGGTCGCCAGCTGACGCGCTGCTGA 1091
|||||: |||: |||: |||: |||: |||: |||: |||: |||
307AlaglyThrSerGlyAlaGlyProGlyThrThrAla..... 318
1090 ACTTGAGAGTGTGTTTTCAGGAATGCGGAGGCTGACGCGGTGATG 1041
319SerSerValGlyValThr 324
1040 GAG.....TATTGTCGGCGGCGCGCCACACCGCGCAACAG 1003
|||||: |||: |||: |||: |||: |||: |||: |||: |||
325 GluThrAlaArgProSerValAlaGlySerGlyThrThrGlyThrValSe 341
1002 CTCTTGGCTGCGGCTTCTTCGATACGGAATCTGATTCGTAGCGTC 953
|||: |||: |||: |||: |||: |||: |||: |||: |||
341 rGlyAlaSerGlySerThrylSerSer..... 350
952 CCAATATATGTCGCGCGCTGTGTAATCGCGCGCTCAATACGCAACG 903
|||: |||: |||: |||: |||: |||: |||: |||: |||
351SerGlySerProGlyAlaThrGlyAlaSerIleGlyGlnPro 364
902 GAAATACAGCGGTGCT.....GCGTCACCAATTCGCGCGGAGTAAT 859
|||: |||: |||: |||: |||: |||: |||: |||: |||
365 GluThrSerArgIleSerValAlaGlySerSerGlyAlaProAlaValSe 381
858 TTGGCATCTTTCGACCCCAACAGGTACGACGAGCGCTGTTGTTGA 809
|||: |||: |||: |||: |||: |||: |||: |||: |||
381 rSerGlyAla..SerGlnAlaAla..GlyThrSerGlnAla..... 393
808 CTTGAGAACCCACCAACCAATCAACGCG.....CTCGGTGTTTC 771

|||||: |||: |||: |||: |||: |||: |||: |||: |||
394ThrThrSerIleGlyAlaGlyThrSerGlyValGlyPhe 408
770 AGACGCGCTGTGC.....AACCAACGTCGATGCGCAATTCACCTTG 727
|||||: |||: |||: |||: |||: |||: |||: |||: |||
408 yStrGlnAlaThrThrPheProGlyGlnAlaGlnGluThrThrArgValGly 424
726 ATATATGATGTCACAAAGCTTTGTTTGACCGACCGCGTC.....AA 683
425 lLeAlaThrGlyThrThrGlyIleValSerArgIleThrLeuGluProG 441
682 TGAATGAATGTCGTCACCTCAACAC.....GGCCGATGCGCGCGCG 636
441 ySerTyrAsnThrGlnAlaThrThrSerIleGlyArg..... 453
635 AATTCATGTGTTTCGATGTTGCGACATTTTCAGACGCGACGTCGCGC 586
453 453
585 AGTCGCTTACACACATGATTTTACGCTCGTCACACG.....GC 545
454 SerGlyThrThrThrAspLeuProGlyGlyThrThrIleValLeuPr 470
544 TCAATACACACANACNTGCTGAAATCNCGCGCTTCTTGATCACA 495
|||||: |||: |||: |||: |||: |||: |||: |||: |||
470 oglyPheSerHisSerSerGln..SerSerIlyProGlySerSerValThr 486
494 ACCACAGGCTGCGCGCAGCGGATGTCATGCA..... 456
487 ThrProGlySerProGlySerGlySerGlyThrGlyThrSerGlyLuph 503
455 .TTGACGAGATGCGGACGCGCTCGGATCGACGCGACGAGATTTTCCTGA 407
|||: |||: |||: |||: |||: |||: |||: |||: |||
503 eSerThrThrValIleSerIleSerSerHisThrGlnAlaThrThrPheI 520
406 ACGACGANTACGACGAGCGAGTCACACAAACCGGATTCGATTCAGTTCNN 357
|||: |||: |||: |||: |||: |||: |||: |||: |||
520 lGlyGlySerGlySerPro.....Gly 527
356 CNAANTCNCGCGCTTAAAGTTGCCACGCTTCGCGCGCGCTGCTTC 307
|||: |||: |||: |||: |||: |||: |||: |||: |||
528 ThrGlySerArgProGlyThrThrGlyGluLeuSerGly.....Th 541
306 GAACCTGATTCGTCGTCGTCCTTCACAGCGCAATTCAGACGAGTGAAGA 257
|||||: |||: |||: |||: |||: |||: |||: |||: |||
541 rThrIleAlaSerGlyAsnAlaThrThrGlnAlaThrThrSerThrGlu 558
256 CGCGCTTTCGCGCGATGATGCGCGATTTTGCTGAACAGCGCGC 207
|||||: |||: |||: |||: |||: |||: |||: |||: |||
558 hrArg.....IleGlyProGlnThrGlyAla 566
206 GTAACACACAGCGCGGATTCCTTGTCTTCAACAGACACTTGCCCTTT 157
|||||: |||: |||: |||: |||: |||: |||: |||: |||
567 GluThrThrValProGlySerGlnValSerGlySerGluThrGlyThrSe 583
156 TTGACGCGCATCGCTTCCTGACTTCATCNAAGGCGCATACGCGCAT 107
|||||: |||: |||: |||: |||: |||: |||: |||: |||
583 rGluAlaValSerAsn.....ProAlaI 591
106 ATTCTTCGCCCAACAGCGACTTCGTAATGACGCGCGCTCA 63
|||||: |||: |||: |||: |||: |||: |||: |||: |||
591 lAlaSerGlySerSerSerThrylThrThrSerGlyAlaSer 605
seq_name: SwissProt_40: FAS1_SCHPO
seq_documentation_block:
ID FAS1_SCHPO STANDARD; PRT; 2073 AA.
AC Q9UG0; P78799;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fatty acid synthase subunit beta (EC 2.3.1.86) [includes: 3-hydroxypalmitoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.61);

DE Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9); [acyl-carrier-protein] acetyltransferase (EC 2.3.1.38); [acyl-carrier-protein] malonyltransferase (EC 2.3.1.39); S-acyl fatty acid synthase thioesterase (EC 3.1.2.14)).

GN FAS1 OR SPAC926.09C.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI_Taxid:4896;

RN [1]

RP SEQUENCE FROM N.A. AND SUBUNIT.

RX MEDLINE-98359895; PubMed-9693066;

RA Niwa H., Katayama E., Yanagida M., Morikawa K.;

RT "Cloning of the fatty acid synthetase beta subunit from fission yeast, coexpression with the alpha subunit, and purification of the intact multifunctional enzyme complex."

RT multifunctional enzyme complex."

RL Protein Expr. Purif. 13:403-413(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-972;

RA Wood V., Rajandream M.A., Barrell B.G., Ramsperger U., Pohl T.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1725-2073 FROM N.A.

RC STRAIN-PR745;

RX MEDLINE-98162722; PubMed-9501991;

RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;

RT "Identification of open reading frames in Schizosaccharomyces pombe cDNAs."

RL DNA Res. 4:363-369(1997).

CC -1 FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THE BETA SUBUNIT CONTAINS DOMAINS FOR: [ACYL-CARRIER PROTEIN] ACETYLTRANSFERASE AND MALONYLTRANSFERASE, S-ACYL FATTY ACID SYNTHASE THIOESTERASE, ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE, AND 3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE.

CC -1 CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH = LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+).

CC -1 CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].

CC -1 CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].

CC -1 CATALYTIC ACTIVITY: (3R)-3-Hydroxypalmitoyl-[acyl-carrier protein] = 2-Hexadecenoyl-[acyl-carrier protein] + H(2)O.

CC -1 CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NAD(+) = trans-2,3-dehydroacyl-[acyl-carrier protein] + NADH.

CC -1 CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-carrier protein] + oleate.

CC -1 SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB010274; BAA36384.1; -

DR EMBL; AL110469; CAB54157.1; -

DR EMBL; D89148; BAA13810.1; -

DR InterPro: IPR001227; Acyltransf.domain.

DR InterPro: IPR003965; Fatty.acid.synth.

DR InterPro: IPR002539; Maoc.dehydratas.

DR Pfam; PF00698; Acyl.transf.1.

DR Pfam; PF01575; Maoc.dehydratas; 1.

DR PRINTS; PR01483; FASYNTHASE.

KW Fatty acid biosynthesis; Multifunctional enzyme; Oxidoreductase;

KW Transferase; Hydrolase; Lyase; NAD; NADP.

FT DOMAIN 1 459 ACETYL TRANSFERASE.

FT 470 858 ENOYL REDUCTASE.

FT DOMAIN 1155 1644 DEHYDRATASE.

FT DOMAIN 1645 2073 MALONYL/PALMITOYL TRANSFERASE.

FT ACT_SITE 270 270 ACETYL TRANSFERASE (BY SIMILARITY).

FT ACT_SITE 1361 1361 DEHYDRATASE (POTENTIAL).

FT ACT_SITE 1828 1828 MALONYLTRANSFERASE (BY SIMILARITY).

FT ACT_SITE 222 222 P -> R (IN REF. 1).

FT CONFLICT 222 222

FT SEQUENCE 2073 AA; 230559 MW; D928270838E7C7C4 CRC64;

align_scores:

Quality: 103.50 Length: 464

Ratio: 0.498 Gaps: 25

Percent Similarity: 44.828 Percent Identity: 20.690

alignment block:

US-09-303-518D-127 x FAS1_SCHPO ..

Align seg 1/1 to: FAS1_SCHPO from: 1 to: 2073

217 TCAGGCAAAATCGCCGATCATCGCGGCAAAAGCGGTACTTCAGTC 266

948 Thcglylsleu.....Metargylleglgluarphethrglns 962

267 GGTG.....GTGATGCCGTGAAGCAACGCAATC..... 300

962 pValglylsthThrleuileglnsnphegluaspleuasnsprpt 979

301GAGTTCGAACGCTACCGCGCCGAGCGTTGGCAAC 336

979 yrProvalalalalargpheluasplalatytrprogluala..... 992

337 TTAAGCGCGANGAANTNNNGCATCGATCGAATCCGGTTGGAC 386

993serThrglnaspleuasnthrglnasplaglnpethery 1006

387 TGGCGTCGATNCCGTCGTCGACAA.....ATCCGCGG 424

1006 rSerleucysersnproheglinsprovalProphelproalal 1023

425 TCGAT..... 432

1023 leaspsprlnrphegluphetyrphelyslsaspsleutrglnser 1039

433 GAGCGTCGCGCATCTTCGCAT.....GCGATGCA 464

1040 Gluaspleualalavalalglguaspsvalglaryvalalale 1056

465 CACCAATCCGTCGCGGACACCGCTGCTGTGATCAAGAACCGGCG 514

1056 uGlnglyPrometcalalalyslnsSerThrlyvalasngluproalal 1073

515 ANGATTTCAAGCANGTNTGCTGTATTGACCGGTTGACGCGTAA 564

1073 ysglu.....leuunspglylleasngluthnls 1083

565 ATC...CATGTGTAGGCA.....GTGCGCGAGAC.....GTGCC 599

1084 lIlelgnlsphelleylslslyspheThryalaglsprlulyslslepr 1100

600 GTGCAAAATGCTGCCAATCGAATCGAATGGAATTCGCGCGCGCATC 649

1100 oile.....Valglutyrrphcglglyvalprop 1110

650 CGGCGGTTTGAGT..... 663

1110 rovalasvalsernlslsSerleuGluservalserValThnGlulu 1126

664 ...GGCAGCAGATTCATTCATTGAGCCGCGGTGCGTCAACAAACCGT 710

1127 AlaglyserlyslsValtyrlyslsleuproglulIlelyserlnsersalale 1143

711 T..... 711

```

1143 uProSerLysLysLeuTrpPheGluLeuAlaGlyProGluTrpThr 1160
712 .....TGACCATC 720
1160 rPheArgAlaIlePheThrTrpGlnArgValAlaLysGlyTrpLysLeu 1176
721 AATTATCAAGATGTAATTGCCATCGACGTTGTTGCACAGCCGCTCT 770
1177 GlnHisAsnProVal.....ArgArgIlePheAlaPro.....Ar 1188
771 GAACACCGAGCCGCGATTCGTTGGGCGTCTCAAGTCAACAACACAC 820
1188 GTyrGlyGlnArgAlaValAlaValLysAspAsnSprThrValValG 1205
821 GCCTCTGGCGTAC.....GTTTGGTGGCGAAGTATGCG 855
1205 LLeuTyrGlyLthrGlnSerGlyAsnTyrValLeuAlaAlaArgLeuSer 1221
856 CAATTTACTGCGGCGCAATTGGTTGACGACGACAAACCGCGATTTCGG 905
1222 .....TyrAspGlyGluThrIleValIleValSerMe 1231
906 TTCGGATTGACGCGCGATTCACACAGCGCCGAC...GATTATTGG 952
1231 rPheGluAsnArgAsnAlaLeuLysLysGlyValHisLeuAspPheLeuP 1248
953 GACGCTAC.....CACAAATCAGATTTCGGTATCGAAGAA 987
1248 helYstYrGluProSerAlaGlyTyrSerProValSerGluIleLeuAsp 1264
988 GCGCCGACGACAAAGAGCTG.....TTCGGCTGGGTT..... 1017
1265 GlyArgAsnAspArgIleLysHisPheTyrTrpAlaLeuTrpPheGlyG 1281
1018 .GCGCCGACGCGGACAAATACATCCATCAGCGCTACGACCGCTCGGCC 1066
1281 uGluProTyrProGluAsnAlaSerIleThrAspThrPheThrGlyProg 1298
1067 TCCTGAAA.....AACAACTCTTCAGTTCACGACGACGCGTCAAC 1107
1298 LuValThrValThrGlyAsnMetIleGlyAspPheCysArgThrValGly 1314
1108 GGTGGCGACCGCGCCATGCGCGATTCGTAATTCAGACGCGTAATGCC 1157
1315 AsnHisAsnGluAlaTyrThrLysArgAlaIleArgLysArgMetAlaPr 1331
1158 GCTAGAC.....ATCC 1168
1331 oMetAspPheAlaIleValIleGlyTrpGlnAlaIleThrLysAlaIleP 1348
1169 TGCCCTACCCCTGTTTGGCGCATTTA.....ATCGTCGCGATACCGAC 1212
1348 heProLysAlaIleAspArgLysPheLeuAlaGlyLeuValHisLeuSer 1364
1213 AGCGCGCAAGCATTTGGTTCGTAATTCAGACGACGAAAGAAC 1254
1365 SerPheArgMetValGlySerHisSerLeuMetGluGlyAsp 1378
seq_name: SwissProt_40:GID_BACSU
seq_documentation_block:
ID GID_BACSU STANDARD; PRT; 435 AA.
AC P39815;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein gld.
GN Gld.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Foulger D., Errington J., 5Kbp of DNA from Bacillus subtilis upstream
RT "Cloning and sequencing of the EMBL/Genbank/DBJ databases.
RT of the codv gene."
RN Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
RP [2]
RC SEQUENCE OF 1-363 FROM N.A.
RC STRAIN-168 / 865;
RA de Jong S.;
RT "Cloning and sequencing of the Top1 gene, the gene encoding B.
RT subtilis DNA topoisomerase I."
RL Submitted (XXX-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AJ000975; CA04423.1; -.
CC EMBL: Z99112; CAB3486.1; -.
CC EMBL: I27797; AAA22764.1; -.
CC Subtilisin, B01.008, gld.
CC Interpro: IPR002218; GIDA.
CC Interpro: IPR000205; NAD_binding.
CC Prodom: PD003738; GIDA.1.
CC PROSITE: PS01280; GIDA.1; FALSE_NEG.
CC PROSITE: PS01281; GIDA.2; 1.
CC Complete proteome.
SQ SEQUENCE 435 AA; 48063 MW; 6C7AB028F484B683 CRC64;

```

```

alignment_scores:
Quality: 103.00 Length: 295
Ratio: 0.824 Gaps: 11
Percent Similarity: 42.373 Percent Identity: 20.678

```

alignment_block:

US-09-303-518D-127 x GID_BACSU ..

Align seg 1/1 to: GID_BACSU from: 1 to: 435

```

331 GCAAACTTAAGCGCGGANGAANTNNGNCATATGATCCGATCCGTTT 380
||||:||||:||||:||||:||||:||||:||||:||||:||||:
11 AlaglyLeuAlaGlySerGluAlaIleAlaIleAlaIleAlaLysArgGly 27
381 GTGACATCGCGCTG...CGTANCCGTCGCGTTCAGCAAAATCCCTGCCG 427
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 eGlnValLysLeuTyrGluMetArgProValLysGlnThrProAlaHis 44
428 ATCCGAGCCGCTGCGCATCTGCTG.....AATCGATGACACCAAT 471
||||:||||:||||:||||:||||:||||:||||:||||:||||:
44 IsthAspLysPheAlaIleGluLeuValCysSerAsnSerLeuAlaGlySer 60
472 CCGCTNCGCGGACGCGCTGTTGTTGATCAAGAAGCCGCGANGATTT 521
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 ThrLeu...AlaAsnAlaValGlyValLeuLysGluGluMet..... 73
522 CAGACGANGTTCCTGCTATGAGCCGTTTGACCGGACGCTAAATGCAATG 571
74 .....ArgAlaLeuAspS 78
572 TGTGTAAGCAGCTGCGCAGAC.....GTGCGCTGTGAAAATGCT 612
::: |||:||||:||||: |||:||||: |||:||||: |||:||||:
78 erAlaIleIleAlaIleAlaAspGlyCysSerValProAlaGlyGlyAla 94
613 GCCAATGCAAAACATGATTTGGCGGCCCG..... 645
::: |||:||||:||||: |||:||||: |||:||||: |||:||||:
95 LeuAlaValAspArgHisGluPheAlaIleAlaIleSerValThrAsnArgVal 111

```

```

646 ....CATCGCGCGGTTTGTAGTGCGACGACATTCATTTCATGAGCCGG 691
      |||||
111 sanh1sPro..... 114
692 TCGGTCAACAACACCGTTTGACCATCAATTCACAGATTAATGGC 741
      |||||
115 .....AsnValThrValIleAsnGluGluValThrGluLeuProGlu 128
742 ATCGGACGTTTGTTCACACGCGCTGTACACCGACGCGGTGATTCG 791
      |||||
129 GlyProThrIleIleIleThrGlyProLeuThrSerGluSer..... 142
792 TTTGGTGTCTTCAGTCACAAACACGCGCTTTCGTCACGTTTGG 841
      |||||
143 .....LeuS 144
842 GTCGCAAAATTCGCAATTA..... 864
      |||||
144 eRLaIleuLysGluLeuThrGlyLysAspTyrLeuTyrPheTyrAsp 160
865 ...GCGGCGAATTTGGTTCACGACGACGCGGTGATTCGCGTTCGT 911
      |||||
161 AlaAlaIleProIleValGluLysAspSerLeuAspMetAspLysVal 177
912 ATTTGAACGCGCGATTACACAAAGCGCGACGATTATTGGGA..... 954
      |||||
177 rLeuLysSerArgTyrAspLysGlyGluAlaIleTyrLeuAsnGlyP 194
955 .....CGCTACCAATCAATCAGATTCCCTTTCGAA 984
194 eThrGluGluLupheAspArgPheHisGluAlaLeuThrSerAlaGlu 210
984 ..... 984
211 ThrValProLeuLysGluPheGluLysGluIlePhePheGluGlyCysMe 227
985 .....GAAGCGCGACGCAAAAGCTGTTCGCT 1012
227 tProIleGluValMetAlaLysArgGlyLysThrMetLeuPheGlyP 244
1013 GGGTTCGCGCGACGCGACAAATCTCCATCAGC 1047
244 rMetLysProValGlyLeuGluHisProValThr 255
seq_name: SwissProt_40:ATPB_DICDH
seq_documentation_block:
ID ATPB_DICDH STANDARD; PRT; 481 AA.
AC P30158;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ATP synthase beta chain (EC 3.6.3.14).
GN ATPB.
OS Dictyofa dichotoma.
OC Eukaryota; Stramenopiles; Phaeophyceae; Dictyotales; Dictyotaceae;
OC Dictyofa.
OC NCBI_TaxID=2876;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92322956; PubMed=1535802;
RA Leitsch C.E.W., Kowalik K.V.;
RT "Nucleotide sequence and phylogenetic implication of the ATPase
RT subunits beta and epsilon encoded in the chloroplast genome of the
RT brown alga Dictyofa dichotoma."
RL Plant Mol. Biol. 19:289-298 (1992).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).

```

```

CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X66939; CAA47370.1; -.
CC PIR: S22509; S22509.
CC DR HSSP: P00829; IBMF.
CC DR Mendel: 2225; DICDH; atpB.1.
CC DR InterPro: IPR004100; ATP-synt_ab.N.
CC DR InterPro: IPR000793; ATPase_AB_C.
CC DR InterPro: IPR000194; ATPase_alpha_beta.
CC DR Pfam: PF00006; ATP-synt_ab.1.
CC DR Pfam: PF00306; ATP-synt_ab.C.1.
CC DR Pfam: PF02874; ATP-synt_ab.N.1.
CC DR PROSITE: PS00152; ATPASE_ALPHA_BETA.1.
CC DR ATP synthetase; Chloroplast; Thylakoid; Membrane; CF(1);
CC KW Hydrolyase; ATP-binding; Hydrogen ion transport.
CC NP_BIND 161 168 ATP (BY SIMILARITY).
CC SQ SEQUENCE 481 AA; 52203 MW; 849B15C113FD1F91 CRC64;

```

```

alignment_scores:
Quality: 103.00 Length: 506
Ratio: 0.468 Gaps: 27
Percent Similarity: 43.478 Percent Identity: 22.530

```

alignment_block:

```

us-09-303-518d-127 x ATPB_DICDH ..
Align seg 1/1 to: ATPB_DICDH from: 1 to: 481

```

```

22 CTAACTGCGCCATCGCGGCGACGCGACGATTCATTATGAGCGC 71
      |||||
11 LeuAsnIle.....GlyTyrIleThrGluValIle.....GlyP 22
72 CGTCATTCACGAGTGCCTGCTTGGCAAGATATACCGGTATGGCC 121
      |||||
22 ovalIleAspAlaValPhe.....SerAlaGlyGluLeuP 34
122 CC.....TNGATGAAGTCAGAGAGCGCATGCC..... 150
      |||||
34 rGlySileTyrAsnAlaLeuGluValLysSerLysAspGlyThrIle 50
151 GTCAAAAAGCGCAAGTCTGTTGAAGACAAAAGATCCGGCGTGT 200
      |||||
51 IleCysGluValGluGlnLeuPheAsnAspAsnArgValArgAlaIle 67
201 GTTACC..... 207
67 aMetSerAlaThrAspGlyLeuGlnArgGlyValGluValIleAspThr 84
208 ..GCGCGGTTTCAGCGCAAAATCGCGCCATTCATCGCGCGAA...AAG 252
      |||||
84 InaIleProIleLeuValProValGlyLysAlaThrLeuGlyArgIlePhe 100
101 AsnValLeuGlyGlnThrValAspAsnIleGluIleGlyThrGlyGluS 117
253 GCGCTACTTCAGTCGCTGATTCGCGTTGA.....GGCAACGA 293
      |||||
294 CGAAATGAGATTGCAAGCTACGCGCGCGAAGCGTTGGCAAAATTAAAG 343
      |||||
117 parGluProIleHisArgProAlaPro.....SerPheThr 130

```

```

344 GCGANGAANTNNNGCAATCTGATCCATCCGTTTGTGACGCTG 393
130 spledunhrlyspralalepnehlutrhglylelysalvalasp 146
394 CGTACCCTCCGTTGACGAAAATCCCTGCGTGCATGCCGACCTTGC 443
147 LeuLeuAlaProTyrArgArg.....GlyGlyLysIleG1 158
444 CATCTTCGATGGATGGACACCAATCCGTGCGGACGACCTTGG 493
158 yLeupheglylala.....GlyAlaGlyLysThrV 169
494 TTGTGATCAAGACCGGANGATTCAGACGANGTGTGCTGATTG 543
169 alLeuilemetleuLeuIleasnIleAlaLys..... 180
544 ACCCGTTGACCGACGCTAAATCCATGCTGTGAAGCAGCTGGGAGA 593
181 .....AlaHisGlyGlyAlaSerValPheGlyGlyAlaGlyLuar 194
594 CGTGGCGTGTGAAATGCTGCCACATGAAACACATGATTCGGCGGC 643
194 gThrArgGlyGlyAsnAspLeuTyrMetGluMetLysGlySerGlyAla 211
644 CGCATCCGCGCGTTGAGTGACGACGACATTCATTTCATT..... 684
211 LeasngluserasnleuserGlySerLysValAlaLeuValTyrGlyIn 227
685 .....GACCGCGTGTGCAACAAACCCGTTGG..... 714
228 MetasnGluProGlyAlaArgMetArgValGlyLeuThrAlaLeuTh 244
715 .....ACATCAATTATCAAGATGATTAATGTC...A 742
244 rMetlaGluTyrPheArgAspIleAsnArgLinspalleuLeuPheI 261
743 TCGGACGTTTGTGCAACAGCCCTCGAACACCGCGGCTGATTGCT 792
261 LeaspsnIlePheArgPheValGlnAlaGlySerGlyValSerAlaLeu 277
793 TTGGGT.....GCTTCACAGTCACAAACCCGCTTGGCTGACCGT 836
278 LeuGlyArgMetProSerAlaValGlyTyrGlnProThrLysGlyThG1 294
837 TTGGGTGCGAAGATATGCAATACGCGGCGAATTGTTGACGAG 886
294 uMetGlyAlaLeuGlnGlnArgIleThr..... 303
887 ACAACCGCGTATTCGCGTTGCGATTGAAAGCGCGATACACAAGGC 936
304 .....SerThrThrGlnGlySerIleThrSerIle 313
937 GCGCAGATTAATTGGACGCTACCAACAATGATTCGTTATCGAAGA 986
314 GlnAlaValTyrVal..... 318
987 AGCGCGCAGAAAGAGCTGTCGGTGGGTTGCGCGGACGCGCAAT 1036
319 .....ProAlaAspAspLeu.....ThrAspProAlaPro..... 328
1037 ACTCCATCAGCGTACGACCTCGGCATTCCTGAAAAACAACTCTTC 1086
329 .....AlaThrThrPheAlaHis.....Leu 335
1087 AAGTCAGACAGCGCTCAAGCGTGGCGACGCGGATGCGCATGG 1136
336 AspAlaThrThrValLeuSer.....ArgGlyLeuAlaAlaLysG1 349
1137 TACTTACGAGCGGCTAATGCGCTAGACATCTGCGCTGCTTGGC 1186
349 yIleTyrProAlaValAspProLeuAspSerThrSerThr...MetLeuG 365

```

```

1187 GCGATTAATATCGTGGCATAC.....GACAGCGGCA..... 1221
365 lnpvalaIleValGlySerGlnHisTyrAspThrAlaGlnLeuValys 361
1222 .....GCATTGGG 1229
382 LysThrLeuGlnArgTyrLysGluLeuGlnAspIleIleAlaLeuG1 398
1230 TTGCTTGGAATTGGACGACAGAAGAC..... 1254
398 yIleAspGluLeuSerGlnGlnAspArgLeuValAlaAspArgAlaArgL 415
1255 .....CTCGCTTGTGACGCTTCGCTGTC..... 1278
415 yslleGluArgPheLeuSerGlnProPhePheValAlaGluValPheThr 431
1279 .....CCGGGCAAAATAC 1290
432 GlySerProGlyLysTyr 437

seq_name: SwissProt_40:RNFC_ECO57
seq_documentation_block:
ID RNFC_ECO57 STANDARD; PRT; 740 AA.
AC P58324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Electron transport complex protein rnfc.
GN RNFC OR Z2636 OR ECS2338.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Iit, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -!- COFACTOR: Binds 2 4FE-4S clusters (potential).
CC -!- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AF005386; BAG56618.1;
DR EMBL: AF002558; BAB35761.1;
DR PROSITE: PS00196; 4FE4S_FERREDOXIN; 2.
FT Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
KW METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL);
FT CONFLICT 651 651 R -> RKAVEAALAKAKRLPQOQANPEEPVDP
      (IN REF. 2)
SO SEQUENCE 740 AA; 80297 MW; 8BE8A9E22DEC56F8 CRC64;

alignment_scores:
Quality: 103.00 Length: 324
Ratio: 0.665 Gaps: 13
Percent Similarity: 47.840 Percent Identity: 21.914

alignment_block:
US-09-303-518D-127 x RNFC_ECO57 ..
Align seg 1/1 to: RNFC_ECO57 from: 1 to: 740

100 GAAGATATGCGGATGCGCCCTGATGAAAGTCAAGAGGCGATGC 149
    :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
48 LysGlnHisIleGlyAlaGlnIleGlyLeuGlyValSerValGlyAsp 64
150 GCGCAAAAAGGCGCAAGTGTCTT.....GAAGCAAAAAGNATCCGG 193
    |||  |||  |||  |||  |||  |||  |||  |||  |||
64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuPro 80
194 GCGTGTGTATTACCGCCGCTTTCAGCGCAAAATCGCGCCGATC.... 237
    |||  |||  |||  |||  |||  |||  |||  |||  |||
81 .....ValHisAlaProThrSerGlyThrValThrAlaIleAlaPro 94
238 .....CATCGCGCGCAAAAAGCGCGTACTGATGCGCGTGTAT 275
    ||  ||  ::  ||  ||  ||  ||  ||  ||  ||  ||
95 HisSerThrAlaHisProSerAlaLeuAlaGlnLeu...SerValIle11 110
276 TGCCGTTGAAGGCAACGAC..... 294
110 eAspAlaAspGlyGlnAspCysTrpIleProArgAspGlyTrpAlaAsp 127
127 yArgSerArgArgArgGlnGlnIleGlnAlaGlnIleHisGlnPheGly 143
328 TTGGCAACTTAAAGCGCGCAGANAATNNGNCATCTGATCCATCCGG 377
    :::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
144 ValAlaIleGlyLeuGlyGly.....AlaGln 151
378 TTGTGGATCGCGCGCGGTANCGTCGCTTTCAGCAAAATCCGCGCGCG 427
    |||  |||  |||  |||  |||  |||  |||  |||  |||
151 yPheProThrGlyVal.....LysLeuGlnGlyGlyG 162
428 ATGCCAGCGCGTTCGCATCTTGCATGATCGATGACCAATCCGCTN 477
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
162 lAspArgIleGlnThrIleuIleIleAsnAlaIleGlnIleGlnIlePro 178
478 GCGGCGAGCCCTGTGTGTGATCAAGAAGCGCGCAGATTTGACAGCG 527
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
179 lIleThrAlaAspArgLeuMetGlnAspCysAlaAlaGlnValValGln 195
528 ANGTNTGCTGATGAGCGCGTTGACAGCG..... 558
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
195 uGlyIleArgIleLeuAlaHisIleLeuGlnProArgGlnIleLeuIleG 212
559 .....CGTAAATTCATGTTGTATGAGGACGCTGGC 588

```

```

212 lylIleGlnAspAsnLysProGlnAlaIleSerMetLeuArgAlaValLeu 228
589 GCGAGCGTCGCCGCTGTAATAATGCTGCCACATCGAACAACATGATCCG 638
    |||  |||  ::  ::  ::  ::  ::  ::  ::  ::
229 AlaAsp.....SerHisAspIleSerMetArgValIlePr 240
639 CGGCGCCCATCGCGCGCT.....TTGAGTG 664
240 oThrLysTrpProSerGlyGlyAlaLysGlnLeuThrTrpIleLeuThrG 257
665 GCAAGCGCATTCATTTTCATGACCGCGTGGTGCAGAAACCAACCGTTGG 714
    ||  ::  ::  ::  |||  |||  |||  |||  |||  |||
257 lylGlnVal.....ProHisGlyGlyArgSerSerAlaIle 269
715 ACCATCATATATCAAGATGTA.....ATTGCATCGAGCGTTGTT 755
    ::  ::  |||  |||  |||  |||  |||  |||  |||
270 GlyValLeuMetGlnAsnValGlyThrAlaTrpAlaValLysArgAlaVal 286
756 TGCAACAGCGCGCTGTGACACAGCGCGGTATGCTTGGGTGTTCTC 805
    |||  ::  |||  |||  |||  |||  |||  |||  |||
286 lIleAspGlyGlnProIleThrGlnArgValValThrIleuThrGlyAla 303
806 AAGTCAACAAACACAGCGCTTGTGTAACGTTTGGTGGCGAAGATCG 855
    ::  ::  |||  ::  ::  |||  |||  |||  |||
303 lAlaIleAlaArgProGlyAsnValTrpAlaArgLeuGlyThrProValArg 319
856 CAATTT.....ACTGGGCGCAATTGTTGACGACACACACCGCGTAT 899
320 HisLeuLeuAsnAspAlaGlyPheCysProSerAlaAspIleMetValIle 336
900 TTCGCGTTTCGTTATTTGACCGC 921
    |||  |||  |||  |||  |||  |||  |||  |||
336 eMetGlyGlyProLeuMetGly 343

seq_name: SwissProt_40:RNFC_ECOLI

seq_documentation_block:
ID RNFC_ECOLI STANDARD; PRT; 740 AA.
AC P77611:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR B1629.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nasimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei S., Seki Y., Sivasubaram S., Tagami H., Takeuchi T.,
RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- FUNCTION: May be part of a membrane complex involved in electron

```

```

CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FE4S BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000258; AAC74701.1; -
DR EMBL: D90806; BAA15384.1; -
DR EMBL: D90807; BAA15391.1; -
DR EMBL: D90808; BAA15414.1; -
DR HSP: P00198; IFCa.
DR Ecogene: E013935; rnfC.
DR InterPro: IPR001450; 4Fe4S-ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF01512; Complex1_51K; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Inner membrane.
KW Complete proteome.
FT METAL 377 377 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 380 380 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 383 383 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 387 387 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 416 416 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 419 419 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 422 422 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 426 426 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 740 AA; 80171 MW; 381506475CD6E01F CRC64;

alignment_scores:
    Quality: 103.00      Length: 324
    Ratio: 0.665        Gaps: 13
    Percent Similarity: 47.840      Percent Identity: 21.914

alignment_block:
US-09-303-518D-127 x RNF_CECOLI ..
Align seg 1/1 to: RNF_CECOLI from: 1 to: 740

100 GAAGATATGCGGATGCGCCCTGATGAAGTCAAGGAAGCGATGC 149
   :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
48 LysGlnHisIleGlyAlaGluGlyIleuGlyValSerValGlyAsp 64
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
150 CGTCAAAAGAGCGCAAGTGTGTTT.....GAAGCAAAAGNATCCG 193
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
64 sValIeuArgGlyIleProIeuThrArgGlyArgGlyLysMetIeuPro 80
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
194 GGTGTGTTTACCGCGCCGTTGAGCAAAATGCGCGCATC..... 237
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
81 .....ValHisAlaProThrSerGlyThrValThrAlaIleLaPro 94
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
238 .....CATCGCGGAGAAAGCGCGTACTTCAAGTGTGATGAT 275
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
95 HisSerThrAlaHisProSerAlaIleuAlaIleu.....SerValIle 110
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
276 TGCCGTTGAAGCAACGAC..... 294
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
110 eAspAlaAspGlyIleAspCysTyrIleProArgAspGlyTrrAlaAsp 127
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
295 .....GAATCGAGTTCGAAGCTACGCGCCGCAAGCG 327
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
127 yTrArgThrArgSerArgGluIleuIleGlyIleValGlyIleHisGlnPhe 143
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

```

328 TTGGCAACTTAAGCGCGGANGAANTNNGNCATCTCATTCATCCG 377
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
144 ValAlaIleuLeuGlyGly.....AlaG 151
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
378 TTGTGAGCTGCGTGGGTANCCGTCGTCAGCAAAATCCGTCGCG 427
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
151 yPheProThrGlyAla.....LysLeuGlnGlyG 162
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
428 ATGCCAGCGCTTGGCCATCTTGTCAATGCGATGACACCAATCCG 477
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
162 LysAspLysIleGluThrLeuIleIleAsnAlaIleGlyCysGluPro 178
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
478 GCGGCAACCCCTGTGTGTGATCAAGAAGCCGANGANCATTTGACAG 527
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
179 IleThrAlaAspAspArgLeuMetGlnAspCysAlaIleValG 195
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
528 ANGTNTGCTGTATGAGCGGTTGACCGAG..... 558
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
195 uGlyIleArgIleLeuAlaHisIleLeuGlnProArgGluIleuIle 212
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
559 .....CGTAAATCCATGTGTGAAGCAGCGTGC 588
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
212 LysIleGluAspAsnLysProGlnAlaIleSerMetLeuArgAlaVal 228
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
589 GCAGAGCTGCGCGTGTGAATGCTGCCAATGCAACATCAATTCG 638
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
229 AlaAsp.....SerAsnAspIleSerLeuArgValIlePr 240
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
639 CGGCGCCATCCGCGCGGT.....TTGAGTG 664
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
240 oThrLysTyrProSerGlyAlaLysGlnLeuThrTyrIleLeuThr 257
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
665 GCAGCGCATTCATTCATTGACCGCGTGTGCAAAACCAACCGTTGG 714
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
257 LysGlnVal.....ProHisGlyIleArgSerSerAspIle 269
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
715 ACCATCATATTCAGATGTA.....ATTGCATGCGAGCGTGTGTT 755
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
270 GlyAlaIleuMetGlnAsnValGlyThrAlaTyrAlaValLysArg 286
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
756 TGCAACAGCGCGTGTGAACACGAGCGGATGCTTGGGTGCTTC 805
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
286 IleAspGlyIleProIleThrGluArgValIleThrLeuThrGlyAla 303
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
806 AAGTCAACACCAACGCGCTCTGCTACCGTTTGGGTGCGAAAGTATG 855
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
303 IleIleAlaArgProGlyAsnValTrrAlaArgLeuGlyThrProVal 319
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
856 CAATTT.....ACTGCGGCGCAATGTGTGACGACGACACCGCGTAT 899
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
320 HisIleuAsnAspAlaGlyPheCysProSerAlaAspIleMetVal 336
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
900 TTCGCGTTCGATTTGAACGCG 921
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
336 eMetGlyGlyProLeuMetGly 343
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

seq_name: SwissProt_40: MUC1_MESAU

seq_documentation_block:
ID MUC1_MESAU STANDARD; PRT; 676 AA.
AC 060528;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Mucin 1 precursor.
GN MUC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
NCBI_TaxID=10036;
RX [1]

```



```

104 raaatgagatgtrpserargleuargleuval..... 116
105 .....
656 CCGGCGGATGCGGCGCGGAAATCATG.....TGTTCGATGT 616
117 ..AspIysIeuValIeuGluTyrIeuPheTyrIeuCysIleAspTyr 132
615 GCGAGCATTTTCAGACGCGCATCTGGCC..... 586
133 ArgIeuValIeuProIysHisIlePheAlaIArgIeuSerIeuValIeI 149
585 ..ACGTGCTTACACACATGATTTTACGTCGGTCAAGCGGCTCAATACC 537
149 yglCysCysAlaHis.....AlaThrAlaArgTyrA 160
536 AGAANACNTGCTGAATCNTGNGC..... 510
160 sperthrasproasnaValAlaSerPheGlySerValAlaIleTyrTr 176
509 ..GCTTCTTGATCACAACACAGAGGCTCGCCGACGCGGATTTGTGCA 461
176 pValSerIeuPheSerAsnValIeIInsSerIeuAspAspIeuPheSerI 193
460 TGCCA...TTGACGAAGATGGCAACGGCTCGGATCGACGCGAGGAT 414
193 IeAlaCysIeuThrIysIleAlaHisArgIeuMetAsnAlaAlaIa 209
413 TTGCTGACGAGGAGNTACGACGCGCATGCCAACACGCGATGATCG 364
210 IeuIlyAsnGlnAlaThrIArgIysProSerSerPro.....GI 223
363 ATTGCNMCNANTTCNTGCCGCTGAAGTTTGCAACGCTCGGCGCGCT 314
223 nThrIArgIAspSerProIle..... 231
313 AGCGTTGCAACTCGATTTGCTGCTCCCTCAACGCAATCAACGACGCG 264
232 .....LeuThrMetAlaProSerThrProVal..... 240
263 TGAAGACGCGGCTTTTCGCGCGGATGATGCGGCGATTTGCT 219
241 ...SerValGlySerThrProProSerThrProSerValIeuProIeAl 256
218 ....GAACNGGCGCGTAACACACGCGCGGATTCCTTTGTCTCA 173
256 alysgInIeuAlaProMetAsnValCysIysAlaHisIleGlnAlaSerA 273
172 ACAGCACTTGCGCTTTTTCAGCGCATCGCT..... 141
273 sngIInsErArgThrIeuThrAlaSerProProGluIleProIeu 289
140 .....TCCTGACTTCATCNCNGGCGCGCATACC 112
290 MetGluProGlnValIyValAsnProGlnValIeuProGlyArgIeuSe 306
111 GGCATATTTCTGCGCAAGCAACGCACTTCGTAATGAGCGCGCGCAT 62
306 rSerIeuSerIysPro.....ValSerIeuProProThrProSer. 319
61 AATGACTTGCTCGGCTCGCGCGGATGCGGAGGTTTACACT 18
320 .....SerProIysValGlyValIyTyrArgPro 328
seq_name: swissprot_40:RNFC_BUCAI
seq_documentation_block:
ID RNFC_BUCAI STANDARD; PRT; 473 AA.
AC P57215;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Electron transport complex protein rnfc.
GN RNFC OR BU115.

```

```

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matunabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. ABS."
RL Nature 407:81-86(2000).
CC -1- FUNCTION: May be part of a membrane complex involved in electron
CC transport (By similarity).
CC -1- COFACTOR: Binds 2 4Fe-4S clusters (Potential).
CC -1- SUBUNIT: Composed of at least six subunits; rnfA, rnfB, rnfC,
CC rnfD, rnfE and rnfG (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE 4FEAS BACTERIAL-TYPE FERREDOXIN FAMILY.
CC RNFC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; AP001118; BAB12833.1;
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR001949; Complex1_51K.
DR Pfam; PF00037; fer4; 2.
DR Pfam; PF01512; Complex1_51K; 1.
DR ProSITE; PS00198; 4Fe4S_FERREDOXIN; 2.
KW Electron transport; Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 337 337 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 340 340 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 343 343 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 347 347 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 376 376 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 379 379 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 382 382 IRON-SULFUR (4Fe-4S) (POTENTIAL).
FT METAL 386 386 IRON-SULFUR (4Fe-4S) (POTENTIAL).
SQ SEQUENCE 473 AA; 54563 MW; 04B64102F1315AEF CRC64;

alignment_scores:
Quality: 100.50 Length: 462
Ratio: 0.455 Gaps: 22
Percent Similarity: 47.835 Percent Identity: 19.697

alignment_block:
US-09-303-518D-127 x RNFC_BUCAI ..
Align seg 1/1 to: RNFC_BUCAI from: 1 to: 473

100 GAAGAATATGCGCGGATAGCGCCCTNGATGAAATGCAAGAGCGCATGC 149
.....
13 AspaPtyrHisAsnValIys.....LeuArgValIysIleAsnIuLy 27
150 CGTCAAAAAGGCCAA...GTGCTGTTTGAAGACAAAGATCCCGGCG 196
||| .....
27 sValIeuArgIyIleProIeuIlePheSerAspPhe.....AsnY 42
197 TGTGTGTTTACCGCGCCNGTTTCAGCAAAATCGCGCCATCATCGCGGC 246
||| .....
42 alProValHisAlaIArgThrSerCylIeuIleGlnAsnIleCysPheAsn 58
247 GAAGAAGCGCGATTCAGTGGTGTGATTCGCGGTTGAAGC..... 288
.....
59 SeraspSerIleIyIysAsnIleIyIleValIleSerProaspTyrIle 75

```

```

289 .....AACGAGAAATCGATTGCAAGCGTACGGC 319
75 uaspgintprleahgleunpProllelyasptlyrlyrallap 92
320 CGAAGCGGTGGCAACTTAAGCGCGANGAANTNNGNCAATGTGATC 369
92 roglulysleulleystile.....llehts 100
370 CAATCGGTTTGGACCTGCGGTANCCCTCCGTC.....AG 410
101 GlnSerGlyVal...ValGlyLeuGlyGlyGlnPheProSerSerly 116
411 CAATAATCCCTGCCGTGATGCCGAGCGCTTCCCAATCTTGTCAATGCGA 460
116 slrsllellepheserlleasnarglalahsthrleullevalasnalay 133
461 TGGACACCAATCCGCTNGCGGAGACCTGTGGTTGTGATCAAGAACCC 510
133 alcluserGluProTyrIleThrSerAspAsnCysLeuIleTyrAsnHis 149
511 GNGGAGATTTGACAGCANGTMTGCTGATGAGCGCTTGGACGAGCG 560
150 lleserGlnleulleclelycyslysllelleCysTPrleThlysl 166
561 TAAATTCATGTGTGTAGGCAAGCTGGCGGAGAGCTGCCGT..... 603
166 elystrValleulleallelelnGlnLuspAsnleGlnSerlleSerl 183
604 .....GAAATGCTGCCAATCGCAACATGCAACATGATTC 636
183 yslleGlnHisleullelyslAsnlySerleuPhelelylleCysllele 199
637 GGGCGCCGATCCGCGCGT.....TTGAG 662
200 LysLysLysTyrProIleaglSerSerlyValleuVallySerleuTh 216
663 TGGCAGCAGCAAT.....CATTCATGACGCGGTGTCGCA 700
216 rgllyslValleuProHisGlylyslHisSerlleasr...lleGly... 230
701 ACAAAACGTTTGGACCAATATTCAGATGATGATTCGCAACGAGCT 750
231 .....TyrleullePheAsnValAlaThrlePheSerlleleystar 244
751 TTGTTGCAACAGCGCTGTAACAGCGCGGTGATGCTTTGGGTG 800
245 AlaIlelleasnclyLysProleuThrGluargValValThleu... 259
801 TTCTCAATGTCACAAACAGCGCTGTGACCGTTTGGGTGCGAAG 850
260 .....MetSerAspLysAsnleuSerGlyAsnPheTyrValArg 274
851 TATGCGAA.....ATPACGGGGCGAATGTGTCGCGCA 885
274 leGlyThrProIleLysTyrPheleThrSerAsnLysleuysGlnSer 290
886 GACAAACGCGGTGATTCGCGTTCCGATTGAACGCGCGCATTAACAG 935
291 .....PheIleAlaSerValTyrleuGlyGly..... 299
936 CGCGACGATTAATTTGGAGCGCTACCAACAATGATTCCTGATGCAAG 985
300 .....ProPheMetGlyLysLysLleAsnleuSerGlyAsn 310
986 AAGCGCGAGCAAGAAGCTGTGGGTGGCGCGGAGCGCGAGCAAA 1035
310 ..... 310
1036 TACTCATCAGCGGTACGACCTCGGCATTTCTGAAAAACAACACTTT 1085
311 HisSerlleleuLysLysThrAsnSerllePheIleThrHisLysLysG 327
1086 CAAGTTACAGACAGCGGTCAACGGTGGCGAGCGCGCATGTGCCGATG 1135

```

```

327 uLysAsnGluSerlle.....SerGluLysThrCyslleArgCysG 341
1136 GFACTTACGAGCGGTAAATGCCCTAGACATCTCGCTACCCGCTT... 1182
341 LyrCysSerTyrValCysProValAsnleuPheProGlnGlnleTyr 357
1183 .....TTGGCGCATTTAATCGGTGCGCATACGACAGCGCGAGCAAT 1226
358 TrpTyrIleLysAsnLysAsnHisValGlnThrLysLysHisTyrValle 374
1227 GGGTGTCTTGGATTTGAGACAGAAAGACCTCGCTTGTGACGCTGTCT 1276
374 uAspCyslleGlu.....CysLysAlaCysGlnLysValC 386
1277 GCCCGGCG.....AAATAC.....GAATNGGCGCG 1302
386 yspProSerTyrIleProleuValLysTyrPheleleGlnGlnLysAsnle 402
1303 CTGTGCGTAAGTGTGTGAACACCTTGAAGAGGAA 1338
403 LeuLysAsnleThrleuGluAsnArgLysLys 414

seq_name: SwissProt_40:HFCL_HUMAN
seq_documentation_block:
ID HFCL_HUMAN STANDARD; PRT; 2035 AA.
AC P51610;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (VP16 accessory protein) (HFCL) (VCAF)
DE (CEF).
GN HCF1 OR HFCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93327419; PubMed=8392914;
RA Wilson A.C., Iamarco K., Peterson M.G., Herr W.;
RT "The VP16 accessory protein HCF is a family of polypeptides processed
RT from a large precursor protein.";
RL Cell 74:115-125(1993).
RN [2]
RP SEQUENCE OF 65-2035 FROM N.A.
RX MEDLINE=95130107; PubMed=7829097;
RA Fratini A., Faranda S., Redolfi E., Zucchi I., Villa A.,
RA Patrosso M.C., Strina D., Susani L., Vezzoni P.;
RT "Genomic organization of the human VP16 accessory protein, a
RT housekeeping gene (HCF1) mapping to Xq28.";
RL Genomics 23:30-35(1994).
RN [3]
RP PROCESSING.
RX MEDLINE=96033796; PubMed=7590226;
RA Wilson A.C., Peterson M.G., Herr W.;
RT "The HCF repeat is an unusual proteolytic cleavage signal.";
RL Genes Dev. 9:2445-2458(1995).
CC -!- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES.
CC -!- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBERT
CC NONCOVALENTLY, ASSOCIATED.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: A SHORT FORM MAY ARISE BY ALTERNATIVE
CC SPLICING. IN THIS VARIANT THE N- AND THE C-TERMINAL FRAGMENTS FAIL
CC TO ASSOCIATE.

```



```

594 GTCGTGGCGGACG.....TGCCTTACACAAATGAGATTTCG.....550
|||||:|||||:|||||:|||||
1074 ValCysSerAsnProProCysGluThrHisGluThrGlyThrHisAsnTh
558 .....CTCGGTCAAA.....550
|||||
1090 GAlaThrThrAlaThrSerAsnMetAlaGluGlnHisGlyCysSerAsnP
549 .....ACGGCTAAATPCCAGACANACTCG 526
|||||:|||||:|||||:|||||
1107 roProCysGluThrHisGluThrGlyThrHisAsnThrAlaThrThrAl
525 TCGGAATGTCGNCGGCT.....507
|||||
1123 aMetSerSerValGlyAlaAsnHisGlnArgAspAlaArgAlaCysA
506 .....TCTTTGATCACAAACCACAGG 486
|||||:|||||:|||||:|||||
1140 lAlaGluThrProAlaValAlaArgGluSerValAlaThrGlyAlaLeu
485 TCTGCCGCGAAGCGGA.....TT 469
|||||:|||||:|||||:|||||
1157 GluAlaAlaGluThrGlySerLysProGlnCysGlnThrArgGlnThrSerAl
468 GGTGTCATCTCGCATTCACGAAGATGGCGACGGCTCGCATTCGACGGCAG
|||||:|||||:|||||:|||||
1173 aThrSerThrThrMethThrValMetAlaThrGlyAlaProCysSerLac
418 GGAATTTTGCTG.....AACGACGGNTACGC 393
|||||:|||||:|||||:|||||
1190 LyrProLeuGluGlyProSerMetAlaArgLysIleProGluGlyArg.....
392 AGCGACGATCCACAAACCGGATTCGGATTCAGTCCGNNNNANTTCNTGCC
|||||:|||||:|||||:|||||
1205 .....SerProAlaPheValGlnIleuAlaLeuProLeuSerSerLys
342 GCTTAAGTTTGGCCACGAGCTCGGGCGGCGTACGCTTCACATCGATTTCGT
1217 sValAlaGluLeuSer.....SerProSerThrLeuLysA 1227
|||||:|||||:|||||:|||||
292 CGTTGGCT.....TCAACGGCAATCAGACGCGCATGAATACG 255
|||||:|||||:|||||:|||||
1227 sPheProAlaGluArgHisSerHisAlaValSerThrAlaAlaMetThr
254 CGCTTTTCG.....CCGGCATGATGGCGGCGCATTTTGGC 220
|||||:|||||:|||||:|||||
1244 ArgSerSerValGlyAlaGluGlyProArg.....MetAlaProValCysG1
219 T.....GAACAGCGCGCGGTAACACACGACCGCGGATTCGTTTGGCTT
1259 uSerLeuGlnGlyLysSerProSerThrThr.....ValThrV 1272
|||||:|||||:|||||:|||||
175 CAACACGACCTTGGCTTTTTCAGCGGCATCGCTTCCTGACTTTCATC 126
|||||:|||||:|||||:|||||
1272 aThrAlaLeuGlnAlaLeuLeuCysProSerAlaThrValThrGlnValA
125 NAGGGCGCATACCGGCATATCTTTCGCCAAGACGCGCATTCGGTAAT 76
|||||:|||||:|||||:|||||
1289 CysSerAsnProProCysGluThrHisGluThrGlyThrHisAsnThrAl
75 GACG.....GGCCGTCATAAAGATCTCGCGTTCGCCG 39
|||||:|||||:|||||:|||||
1305 aThrThrSerAsnAlaGluSerAlaGlnArgValCysSerAsnProPro
seq_name: SwissProt_40_ODP2_MYCA

```

seq_documentation_block:		
ID	ODP2_MCCA	STANDARD; PRT; 438 AA.
AC	Q49110.	
DT	30-MAY-2000	(Rel. 39, Created)
DT	30-MAY-2000	(Rel. 39, Last sequence update)
DT	30-MAY-2000	(Rel. 39, Last annotation update)

DE	Dihydrodiploamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12) (E2).	
DN	PDHC OR ODP2.	
OS	Mycoplasma capricolum.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;	
OC	Entomoplasmataceae.	
OX	NCBI_Taxid=2095;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97001869; PubMed=8844861;	
RA	Zhu P.P., Peterkofsky A.;	
RT	"Sequence and organization of genes encoding enzymes involved in	
RT	pyruvate metabolism in Mycoplasma capricolum.";	
RL	Protein Sci. 5:1719-1736(1996).	
CC	-1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL	
CC	CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE	
CC	COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),	
CC	DIIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPAMIDE DEHYDROGENASE	
CC	(E3) (BY SIMILARITY).	
CC	-1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrodiploamide = CoA + S-	
CC	acetyl-dihydrodiploamide.	
CC	-1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL	
CC	COFACTOR (BY SIMILARITY).	
CC	-1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL	
CC	SYMMETRY (BY SIMILARITY).	
CC	-1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.	
CC	-1- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: U62057; AAC44344.1; -.	
DR	HSSP: P11961; 1B55.	
DR	InterPro: IPR001078; 2oxoacid_dh.	
DR	InterPro: IPR000089; Biotin_lipoyl.	
DR	InterPro: IPR003016; Lipoyl.	
DR	InterPro: IPR004167; e3_binding.	
DR	Pfam: PF00198; 2-oxoacid_dh; 1.	
DR	Pfam: PF00364; biotin_lipoyl; 1.	
DR	Pfam: PF02817; e3_binding; 1.	
DR	ProDom: PD001115; 2oxoacid_dh; 1.	
DR	ProSITE: PS00189; LIPOYL; 1.	
KW	Glycolysis; Transferase; Acyltransferase; Lipoyl.	
FT	BINDING 42 42 POTENTIAL	
FT	ACT_SITE 411 411	
SO	SEQUENCE 438 AA; 46927 MW; 4BF83B69/480B4AB CRC64;	

```

alignment_scores:      Quality: 100.00      Length: 446
                       Ratio: 0.508      Gaps: 23
Percent Similarity:    44.170      Percent Identity: 19.058

alignment_block:
US-09-303-51BD-127 x ODP2_MYCCA ..

Align seg 1/1 to: ODP2_MYCCA from: 1 to: 438

133 GTCAAGAGAAGCGGATGCCGTCAAAAAGCCACGAAGCTGT.....GA 176
||||| |||||:::|||||:::||||| |||||
23 ValLysValGIyAspValIvalValylglnSerLeuYrpehValGI 39
|||
177 AGACAAAAAAGNATTCGGGGCGTGGTGTTAACGCCCCNGATTACAGCAA 226
||| ::| |||||:::|||||
39 uhrasplysValalnsergluite...ProAlaProValalabelylsi 55
|||||:::|||||:::|||||
227 TGCCGCACATTCAT.....CGCGGCAAAAGCGCCTACTTCAGTCGTC 270
|||||:::|||||:::|||||

```

```

55 leaIaValIleAsnIleLysAlaGlyGlnGluIleLysValAlGlyAspVal 71
271 GTGATTGGCGTTGAA...GGCAGCAGCAATTCAGTTCGAAACCGTACGC 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 ValMetGluIleGluAspGlySerAspHisSerAlaThrSerGlu..... 86
318 GCCCGAACCGCTGGCAAACTTAAGCGCGGANGANTNNGNCAATCGA 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 ProLysAla..... 89
368 TCCATTCGCGTTTGGAGCTGCGCTGACCTGACCTGCGTTCAGCAAAATC 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 .....GluThrLysSerGluAlaLysVal 97
418 CCTGCGCTGATGCCGAGCGCTGGCCATCTTCGTCATGCGATGGAGACAC 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 GluValValAlGluGluAsn.....AlaSerValValAlGlyAl 109
468 CAATCCGCTNCGGCGAGACCTGTGGTGTGATCAAGAACCCGCGANG 517
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 aThrProValSerAsnAsp.....ValIleValAlGlyGlnThrThr 124
518 ATTTCAGACGANGTGTGCTGATTGACCGCTTGACCGAGCGTAAATC 567
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 hrValAsnLysSerSerThrIleLysAlaThrProLeuAlaArgLysVal 140
568 CATGTGTGTAGCAGCTGCGCGCAGACGTG..... 597
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 .....AlaAlaIleAspLeuAsnIleAspLeuSerLeuValThrPr 153
598 .....CCGTCTGAAT.....GCTGCCACATCGAAGAACACATGAA 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 oThrGlyProAsnGlnArgIleLeuValAlaAspIleLysAsnHisGln 170
634 .....TTGGCGCGCCCGCAGCCGCGCC 654
170 laSerSerThrGlnLeuAlaSerGlnProIleSerGlnProIleProThr 186
655 GGTGTGAGTGGCAGCGAC.....ATTCAATTCATTGA 686
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 ProSerProSerAlaHisGlnThrIleAlaProThrIleLysValAlG 203
687 GCCCGTGGT..... 696
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 uProSerAlaProLeuSerTrpAspGluValProMetAsnGlyValArgL 220
697 ..GCAACAAACCGTTTGGACATCAATTAATCAAGATGAATTCGATC 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 yAlaIleThrValLysAlaMetThrLysSerHisThrGluIleAlaAla 236
745 GGACGTTTGTTCACAGCGCGCTGACAGCAGCGCGCTGATTCGCTT 794
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 .....ThrGlyMetLysAsnThrAsp..... 243
795 GGGTGTTCATCACTCAACAAACCGCTTTCGATCCGTTTG... 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 .....IleThrGlnThrHisLysMetArgThrGluLeuLysA 256
841 .....GGTGGCAAGTATCGCAAT..... 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 spHisAlaAlaIleSerGlyIleLysLeuThrTrpLeuAlaPheIle 272
862 .....ACTGGGGCGAATGTGTGACGACAGACACCGCGTGAATTCGCG 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 LysAlaValAlaLysSerLeuArgAspMetProAsnIleAsnValAlG 289
906 TTCGTTATTCAGCGCGCGAT..... 927
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 yAspPheAlaAsnAsnLysIleGlnPheMetHisAsnIleAsnIleGly 306
928 .....ACACAAGCGCG 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 leaIaValaAspThrProAsnGlyLeuMetValProValIleLysGlyAla 322

```

```

940 CAGCATTTATTTGGAGCGCTACCAATCAATCATTTCCGTTATCGAAGACG 989
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 ...AspHisLeuSerValPheGlnIleAlaIleLysIleSerGlu..... 336
990 CCCGACCAAGACGTGTGCGGTGGTGGCGCGCGCGACCAATACT 1039
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 .....LeuAlaAsnLysAlaLysAspGlyL 345
1040 CCATCACGCGTACG.....ACCGTGGCGCATTTTC 1068
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 yLeuThrArgAlaGluMetThrGluAlaThrPheThrValSerAsn 361
1069 CTGAACAAACAACTCTCAAGTTCACGACAGCGCGTCAACGCGCGACG 1118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 GlySerValGlyLeuAspTyrAlaThrProIleIleAsnSerProGlu 378
1119 CGCCATGTCGCGCATGTGACTTACGAGCG..... 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 rAlaIleLeuGlyValGlyThrMetSerGlnThrProLeuTyrIleAsn 395
1150 .....GTATGCCGCTAGACATC 1167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 yGluLeuGlnLysArgPheIleMetProLeuSerMet 407

seq_name: swissprot_40:PMAL_DUNAC
seq_documentation_block:
ID PMAL_DUNAC STANDARD; PRT; 1103 AA.
AC P54210;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plasma membrane ATPase (EC 3.6.3.6) (Proton pump).
GN DHAL.
OS Dunalipella acidophila.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Dunalipellaceae; Dunalipella.
OX NCBI_TaxID=38272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAG 19.85;
RX MEDLINE=97127854; PubMed=8972605;
RA Weiss M., Pick U.;
RT "Primary structure and effect of pH on the expression of the plasma
RT membrane H(+)-ATPase from Dunalipella acidophila and Dunalipella
RT salina."
RL Plant Physiol. 112:1693-1702(1996).
CC -! FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYNPORT. THE RESULTING
CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIANE
CC GROWTH RESPONSES.
CC -! CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(IN) = ADP + PHOSPHATE +
CC H(+)(OUT).
CC -! SUBCELLULAR LOCATION: Integral membrane protein.
CC -! SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY IT1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54690; AAB49042.1; -.
CC InterPro: IPR004014; Cation_ATPase.
CC InterPro: IPR001757; E1-E2_ATPase.
CC InterPro: IPR000695; HATPase.
CC InterPro: IPR001454; Hydrolase.
CC Pfam: PF00690; Cation_ATPase_N; 1.

```



```

seq_name: SwissProt_40:YGI_F_YEAST
seq_documentation_block:
ID YGI_F_YEAST STANDARD; PRT; 551 AA.
AC P53214;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 57.5 kDa protein in YMA7-RPS25A intergenic region.
GN YGI023W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=8288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
   chromosome VII."
RL Yeast 13:1077-1090(1997).
CC -1- SIMILARITY: TO YEAST MID2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z72807; CA97006.1; -.
DR SGD: S0003255; MTU1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 104 107 POLY-SER.
FT DOMAIN 111 120 POLY-SER.
FT DOMAIN 124 138 POLY-SER.
FT DOMAIN 142 146 POLY-SER.
FT DOMAIN 182 188 POLY-SER.
FT DOMAIN 215 218 POLY-SER.
FT DOMAIN 222 232 POLY-SER.
FT DOMAIN 236 264 POLY-SER.
FT DOMAIN 269 275 POLY-SER.
FT DOMAIN 278 281 POLY-SER.
FT DOMAIN 284 291 POLY-SER.
FT TRANSMEM 362 382 POTENTIAL.
FT DOMAIN 468 474 POLY-ASP.
SQ SEQUENCE 551 AA; 57527 MW; 86D94341B319951 CRC64;

alignment_scores:
Quality: 99.50 Length: 199
Ratio: 0.865 Gaps: 7
Percent Similarity: 57.789 Percent Identity: 24.623

alignment_block:
US-09-303-518D-127/rev x YGI_F_YEAST ..
Align seg 1/1 to: YGI_F_YEAST from: 1 to: 551
635 AATCATGTGTTTGCATGTGTCAGACGATTTTCAGACGCGACGCTGCGGCC 586
      ::||| ::||| ::||| ::|||
128 SerSerSerIleSerProSerSerSerSerThrIleIleSerSerSe 144
585 AGTCGCTTACACACATGATTTTACGCTCGCTC..... 552
      ::||| ::||| ::||| ::|||
144 rSerSerLeuProThrPheThrValAlaSerThrSerSerThrValAla 161
551 ..AAAGGCTCAATACACAGACGATTCGTCGAATCTGCGCGGCTTCT 504
      ::||| ::||| ::||| ::|||
161 erSerThrLeuSerThrSerSerSerLeuValIleSerThrSerSer 177

```

```

503 TTGATCAACACACAGGCTGCGGACGAGGATGTGTCATCGCATT 454
      ::||| ::||| ::||| ::|||
178 ThrPheThrPheSerSerSerSerSerSerSerIleSerSerIle 194
453 GACGAGATGCGGACGCTCGCATCGACGCGGAGGATTTGCTGAAG 404
      ::||| ::||| ::||| ::|||
194 eSerThr.....SerValSerThr..... 200
403 GACGAGTACCGACGCGCATCCAAACCGGATGATGACATTGCNNCN 354
      ::||| ::||| ::||| ::|||
201 .....SerSerValTyValProSer.....SerSer 209
353 ANTTCNTGCGCGCTTAAGTTGCAACGCTTGCGGCGGAGCTGCGAA 304
      ::||| ::||| ::||| ::|||
210 ThrSerSerPro.....ProSerSerSerGluLeuThrSerSe 223
303 CTGATTTGCTGCTTGCTTCAACGCGCATCACGACGATGAATACG 254
      ::||| ::||| ::||| ::|||
223 rSerTySerSerSerSerSerSerSerThrLeuPheSerTySerSer 240
253 GCTTTTGGCCGCGCATGATGCGCGCATTTTGCTGAACGCGCGGTA 204
      ::||| ::||| ::||| ::|||
240 erPheSer.....SerSerSerSerSerSerSerSerSer 252
203 AACACACAGCGCGGATNCTTTTGTCTTCAACAGCAGCTTGCTTTT 154
      ::||| ::||| ::||| ::|||
253 SerSerSerSerSerSerSerSerSerSerSerTyThrPheThrLeu 269
153 GACGCGATCGCTTCTTCACTTCATCNAAGGCGCATACCGGCAAT 105
      ::||| ::||| ::||| ::|||
269 rThrSerSerSerSerSerSerSerSerSerSerTyProSerPheS 286
104 ....TCTTGGCCACACGACGCTTGGTATGACGGGCGCGTGA 63
286 erSerSerSerSerSerSerSerSerSerSerSerThrSerThrSer 301

seq_name: SwissProt_40:HFC1_MESAU
seq_documentation_block:
ID HFC1_MESAU STANDARD; PRT; 2090 AA.
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Host cell factor C1 (HCF) (Vp16 accessory protein) (HFC1) (VCAR)
   (HCF).
GN HFC1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN 1;
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RT Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSCRIPTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -1- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPGC/THET, WITHIN THE HCF REPEAT (BY

```

